

**UNIVERSIDAD COMPLUTENSE DE MADRID**  
**FACULTAD DE CIENCIAS QUÍMICAS**  
**DEPARTAMENTO DE BIOQUÍMICA Y BIOLOGÍA MOLECULAR**



**TESIS DOCTORAL**

**Estudio de splicing alternativo y variantes sin clasificar en  
genes de susceptibilidad al cáncer de mama y ovario:  
implicaciones clínicas**

MEMORIA PARA OPTAR AL GRADO DE DOCTORA

PRESENTADA POR

**Irene López Perolio**

DIRECTOR

**Miguel de la Hoya Mantecón**

**Madrid, 2018**

**UNIVERSIDAD COMPLUTENSE DE MADRID**

FACULTAD DE CIENCIAS QUÍMICAS  
Departamento de Bioquímica y Biología Molecular



**TESIS DOCTORAL**

***ESTUDIO DE SPLICING ALTERNATIVO Y  
VARIANTES SIN CLASIFICAR EN GENES DE  
SUSCEPTIBILIDAD AL CÁNCER DE MAMA Y  
OVARIO: IMPLICACIONES CLÍNICAS***

Memoria para optar al Grado de Doctor Presentada por  
**IRENE LÓPEZ PEROLIO**

Director  
**MIGUEL DE LA HOYA MANTECÓN**

**Madrid, 2018**







Don Miguel de la Hoya Mantecón, Doctor en Ciencias e Investigador I3-SNS-A del Instituto de Investigaciones Sanitarias San Carlos (IdISSC)

**CERTIFICA:**

Que el trabajo “ESTUDIO DE SPLICING ALTERNATIVO Y VARIANTES SIN CLASIFICAR EN GENES DE SUSCEPTIBILIDAD AL CÁNCER DE MAMA Y OVARIO: IMPLICACIONES CLÍNICAS” que presenta Dña. Irene López Perolio ha sido realizado bajo su dirección y reúne, a su juicio, todos los requisitos necesarios para que sea presentado ante el tribunal correspondiente y pueda optar al Grado de Doctor en Bioquímica, Biología Molecular y Biomedicina por la Universidad Complutense de Madrid.

Para que así conste, y a los efectos oportunos, firma la presente en Madrid a 02 de Octubre de 2017.

Dr. Miguel de la Hoya Mantecón



Esta investigación ha sido posible gracias a la financiación aportada por los proyectos de investigación

- Proyecto de Investigación del Instituto de Salud Carlos III PI12/00539 (supported by European Regional Development FEDER Funds).
- Proyecto de Investigación del Instituto de Salud Carlos III PI15/00059 (supported by European Regional Development FEDER Funds).
- *European Union's Horizon 2020 research and innovation programme* under grant agreement No 634935(BRIDGES).





Parece ayer cuando comencé mi aventura en el laboratorio de Oncología Molecular del Hospital Clínico San Carlos, y sin embargo, han pasado cuatro años desde que aterricé con el propósito de realizar mi proyecto de máster.

Estos cuatro años he tenido la oportunidad de crecer a nivel personal y profesional rodeada de personas excelentes, y con las que comparto esta ilusión por la investigación. Echo la vista atrás y la verdad que no me importaría volver a comenzar dicha etapa, ya que he disfrutado de cada momento. Desde las colas para poder usar el equipo de electroforesis capilar, que estaba reservado porque había que poner MLPAs y secuencias urgentes, hasta los momentos de limpieza y organización. Y por supuesto, nuestra hora del café.

Esta tesis doctoral no hubiera sido posible sin la ayuda de mi director de tesis, el Dr. Miguel de la Hoya. Quiero agradecerte la confianza depositada en mí para la realización de este proyecto. Gracias a tu confianza y enseñanza, he podido comenzar a alcanzar mi sueño de convertirme en investigadora. Aún me queda un largo camino que recorrer, pero sé que con los conocimientos y competencias que he adquirido trabajando contigo, lograré alcanzar mi sueño.

Además, quiero dar las gracias a la Dra. Trinidad Caldés y al Profesor Eduardo Díaz Rubio por acogerme en el laboratorio de Oncología Molecular. Quiero agradecerles la oportunidad que me han brindado al permitirme empezar aquí mi carrera de investigación. Y quiero destacar el apoyo diario de la Dra. Trinidad Caldés en la realización de este proyecto.

También quiero agradecer a todo el personal del laboratorio de Oncología Molecular, tanto a los que están como a los que han pasado por él, el cariño y apoyo que me habéis brindado durante todos estos años. Sin vosotros, este proyecto no hubiera sido igual. He aprendido mucho de cada uno de vosotros, y de verdad que me quedo con un buen recuerdo de todos estos años. Quiero darle las gracias a Marilo, porque sin sus cafés mañaneros, el comienzo del día habría sido muy duro. Gracias, porque el tiempo que pude trabajar codo con codo contigo, pude aprender mucho. Gracias a Patricia, por sus buenos consejos, y sus ánimos en el día a día, y por nuestras prácticas de inglés. Espero en el futuro poder echarte esos partidos de padel que hemos hablado muchas veces.

Quiero agradecerte Alicia, por echarme una mano en la cuantificación de muestras, y haberme enseñado a usar el MagnaPure. Y quiero agradecerte que siempre tengas palabras de ánimo.

Paula, gracias a tu orden y organización, no me ha faltado de nada para poder realizar los experimentos.

Vanessa, he podido aprender mucho de ti, tanto a nivel profesional, como a nivel personal. Además, he podido descubrir sitios muy interesantes para bucear. Pilar, gracias por tu ayuda, ya no sólo en la realización de esta tesis, sino por los consejos que he recibido por tu parte también de cara al futuro. Gracias por tu ayuda. También Inma y Beatriz, daros las gracias por vuestro apoyo todo este tiempo. Y a todos mis compañeros Pre-doctorales, en especial a Mateo. Gracias por hacer más deliciosos los desayunos con tus tartas y magdalenas. Espero que todos lleguéis a terminar la tesis doctoral, y podáis seguir en esta aventura de la investigación.

Quiero daros las gracias a todos vosotros, porque si el día de mañana me marcho a otro lugar a continuar con mi aventura, echaré de menos a esta pequeña familia que todos hemos creado. Además, quiero darles las gracias a todo el servicio de la consulta de consejo genético, a Pedro e Isabel por su ayuda tanto en la mediación con los pacientes, como en la recolección de muestras de RNA.

También quiero dedicarles esta tesis doctoral a mis padres. Porque sin su esfuerzo y sacrificio por brindarme la mejor educación y formación, no hubiera conseguido llegar hasta donde estoy. Quiero agradecerles, junto con mi hermano, por animarme a no abandonar al comienzo de la tesis. Reconozco que estuve a punto de abandonar este camino al ver que mis intentos por conseguir financiación se veían frustrados una y otra vez, y vosotros me supisteis animar y aconsejar a que tuviera paciencia, que al final todo termina saliendo. Y es verdad, de una u otra manera, todo sale. He conseguido terminar la tesis, que muchos han calificado como “la obra del Escorial”.

Quiero agradecerle a mi novio, por apoyarme en cada decisión que he tomado. Por apoyarme a iniciar esta loca aventura de dedicarme a la investigación, aun sabiendo que en algún momento puede que tenga que realizar estancias en el extranjero, y eso implique tener que estar separados un tiempo. Gracias, porque a pesar de que nuestros trabajos son diferentes, siempre hemos tratado de afrontar los problemas como equipo. Quiero darte las gracias por saberme animar, y sacarme una sonrisa aun cuando las cosas se ponen difíciles.

Y también quiero dar las gracias a mis amigos, por vuestro ánimo. De entre vosotros quiero mencionar a Elena. Quiero darte las gracias porque has sido un importante apoyo tanto a nivel profesional, como a nivel personal cuando Mario estuvo fuera. Lucía también quiero

darle a ti las gracias porque siempre has tenido un hueco para verme y hablar conmigo cuando lo he necesitado.

Quiero daros las gracias a todos, por apoyarme y animarme a alcanzar parte de mi sueño. Gracias.



A todos los que habéis creído en mí, y me habéis ayudado a lograr este reto.



# Índice





# ÍNDICE

---

ÍNDICE.....	17
ÍNDICE DE FIGURAS .....	21
ÍNDICE DE TABLAS .....	25
RESUMEN.....	27
SUMMARY .....	31
GLOSARIO DE ABREVIATURAS .....	33
INTRODUCCIÓN.....	37
<b>1. Susceptibilidad genética al cáncer de mama y ovario: <i>BRCA1</i>, <i>BRCA2</i>, <i>PALB2</i>, <i>RAD51C</i> y <i>RAD51D</i> en contexto .....</b>	<b>39</b>
1.1. El Cáncer es una enfermedad genética. El modelo CGE .....	39
1.2. Susceptibilidad genética al cáncer .....	40
1.3. Heredabilidad y arquitectura genética subyacente .....	41
1.4. El modelo poligénico de susceptibilidad al cáncer de mama y ovario .....	44
<b>2. <i>BRCA1</i>, <i>BRCA2</i>, <i>PALB2</i>, <i>RAD51C</i> y <i>RAD51D</i> cooperan en la regulación de la homeostasis del DNA de doble hebra .....</b>	<b>47</b>
2.1. <i>BRCA1</i> .....	47
2.2. <i>BRCA2</i> .....	48
2.3. <i>PALB2</i> .....	49
2.4. <i>RAD51C</i> y <i>RAD51D</i> .....	50
2.5. Muchos de los genes implicados en susceptibilidad genética al cáncer de mama y ovario, incluyendo <i>BRCA1</i> , <i>BRCA2</i> , <i>PALB2</i> , <i>RAD51C</i> y <i>RAD51D</i> , participan en la denominada ruta FA/BRCA, esencial para la reparación por recombinación homóloga de distintas lesiones en el DNA .....	52
<b>3. Susceptibilidad genética al cáncer de mama y ovario en la clínica: de <i>BRCA1</i> y <i>BRCA2</i> a los paneles genéticos.....</b>	<b>54</b>
3.1. Test genético en cáncer de mama y ovario familiar .....	54
3.2. El problema de las variantes genéticas de significado clínico incierto.....	57
3.2.1. Clasificación Cualitativa.....	60
3.2.2. Clasificación Cuantitativa .....	62
3.3. Un análisis de splicing puede aportar información importante para la clasificación clínica de variantes genéticas en <i>BRCA1</i> y <i>BRCA2</i> .....	63
3.4. Se ha demostrado que el splicing alternativo proporciona información potencialmente útil para la clasificación clínica de variantes genéticas en <i>BRCA1</i> y <i>BRCA2</i> .....	67
OBJETIVOS .....	71
MATERIALES Y MÉTODOS.....	75
<b>1. Obtención de RNAs y síntesis de DNA complementario.....</b>	<b>77</b>
1.1. RNA de líneas celulares linfoblastoides.....	77
1.2. RNA obtenido de células nucleadas de sangre periférica .....	77
1.3. RNA de tejidos de Mama y Ovario humanos normales .....	77
<b>2. Cuantificación de RNA.....</b>	<b>78</b>
<b>3. Síntesis de DNA complementario .....</b>	<b>78</b>
A. <u>Estudios <i>in vitro</i> de splicing</u> .....	78

1. Selección de mensajeros: transcrito de referencia .....	78
2. Diseño de cebadores .....	79
3. PCR .....	79
4. Electroforesis en gel de Agarosa .....	80
5. Análisis de fragmentos por electroforesis capilar .....	80
6. Secuenciación Sanger .....	81
B. Caracterización del <i>splicing</i> alternativo .....	82
1. RNAseq .....	82
2. Designación de eventos de <i>splicing</i> alternativo, anotación biológica (biotipo) y anotación funcional .....	83
3. Análisis cuantitativo del <i>splicing</i> alternativo mediante PCR digital .....	84
3.1. Diseño de ensayos TaqMan .....	84
3.2. Validación Ensayos TaqMan .....	85
3.3. PCR digital .....	85
RESULTADOS .....	87
I. <u>Análisis de <i>splicing</i> y clasificación clínica de variantes genéticas en <i>BRCA1</i> y <i>BRCA2</i></u> .....	89
II. <u>Caracterización del <i>splicing</i> alternativo en <i>RAD51C</i>, <i>RAD51D</i> y <i>PALB2</i></u> .....	101
1. Anotación de eventos .....	101
2. Caracterización del <i>splicing</i> alternativo en el gen <i>RAD51C</i> .....	105
2.1. Ensayos RT-PCR/EC realizados en el locus <i>RAD51C</i> .....	110
3. Caracterización del <i>splicing</i> alternativo en el gen de susceptibilidad <i>RAD51D</i> .....	113
3.1. Identificación del exón 3 de referencia .....	113
3.2. Anotación de eventos de <i>splicing</i> alternativo .....	116
3.3. Ensayos RT-PCR/EC realizados en el locus <i>RAD51D</i> .....	120
3.4. Contribución del evento de <i>splicing</i> alternativo $\Delta 3_5$ sobre la expresión global de <i>RAD51D</i> .....	123
4. Caracterización del <i>splicing</i> alternativo en el gen de susceptibilidad <i>PALB2</i> .....	125
4.1. Electroferogramas obtenidos a partir de cada RT-PCR en la caracterización de <i>PALB2</i> .....	128
4.2. Contribución de $\Delta 11,12$ a la expresión global de <i>PALB2</i> .....	133
DISCUSIÓN .....	135
1. Impacto del <i>splicing</i> alternativo en la clasificación clínica de variantes genéticas en <i>BRCA1</i> y <i>BRCA2</i> .....	137
1.1. Análisis de <i>splicing</i> en variantes genéticas sinónimas en <i>BRCA1</i> y <i>BRCA2</i> .....	137
1.2. Análisis de <i>splicing</i> en variantes genéticas <i>missense</i> en <i>BRCA1</i> y <i>BRCA2</i> .....	139
1.3. Análisis de <i>splicing</i> en variantes genéticas intrónicas en <i>BRCA1</i> y <i>BRCA2</i> .....	140
1.4. Análisis de <i>splicing</i> en variantes Clase-1 multifactorial .....	143
2. Impacto del <i>splicing</i> alternativo en la clasificación clínica de variantes genéticas en <i>RAD51C</i> , <i>RAD51D</i> y <i>PALB2</i> .....	147
2.1. Clasificación clínica inicial de variantes genéticas en <i>RAD51C</i> , <i>RAD51D</i> y <i>PALB2</i> .....	147
2.1.1. <i>RAD51C</i> .....	148
2.1.2. <i>RAD51D</i> .....	151
2.1.3. <i>PALB2</i> .....	152
2.2. <i>Splicing</i> alternativo y Diseño experimental de estudios <i>in vitro</i> de <i>splicing</i> en <i>RAD51C</i> , <i>RAD51D</i> y <i>PALB2</i> .....	155

CONCLUSIONES .....	159
ANEXO I .....	163
1. Selección de Transcritos de Referencia para la anotación de eventos de <i>splicing</i> alternativo .....	165
1.1. Cebadores en el locus <i>BRCA1</i> .....	167
1.2. Cebadores en el locus <i>BRCA2</i> .....	168
1.3. Cebadores en el locus <i>RAD51C</i> .....	168
1.4. Cebadores en el locus <i>RAD51D</i> .....	169
1.5. Cebadores en el locus <i>PALB2</i> .....	169
2. Análisis cuantitativo del <i>splicing</i> alternativo mediante PCR digital .....	170
ANEXO II .....	175
1. Análisis de <i>splicing</i> alternativo en portadores de variantes genéticas en <i>BRCA1</i> .....	175
2. Análisis de <i>splicing</i> alternativo en portadores de variantes genéticas en <i>BRCA2</i> .....	180
REFERENCIAS .....	189
ARTÍCULOS PUBLICADOS O EN PREPARACIÓN .....	203
PARTICIPACIÓN EN CONGRESOS .....	203



# ÍNDICE DE FIGURAS

---

## FIGURAS INTRODUCCIÓN

<b>Figura 1:</b> Loci de susceptibilidad al cáncer de mama según su frecuencia y riesgo relativo.....	42
<b>Figura 2:</b> Gráficos del porcentaje de heredabilidad para los loci de susceptibilidad al cáncer de mama y ovario .....	43
<b>Figura 3:</b> Gráficos del riesgo acumulado de desarrollar cáncer de mama y ovario en mujeres portadoras de mutaciones patogénicas en <i>BRCA1</i> .....	46
<b>Figura 4:</b> Esquema de la proteína BRCA1.....	48
<b>Figura 5:</b> Esquema de la proteína BRCA2.....	49
<b>Figura 6:</b> Esquema de <i>PALB2</i> .....	50
<b>Figura 7:</b> Familia de genes <i>recA/RAD51</i> .....	51
<b>Figura 8:</b> Esquema de la ruta FA/BRCA .....	53
<b>Figura 9:</b> Estructura de los límites intrón-exón.....	64
<b>Figura 10:</b> Secuencias reguladoras que median el proceso de <i>splicing</i> .....	64
<b>Figura 11:</b> Análisis <i>in vitro</i> realizado por RT-PCR/EC.....	69

## FIGURAS MATERIALES Y MÉTODOS

---

<b>Figura 12:</b> Calidad de la carga y separación de los clusters .....	86
--	----

## FIGURAS RESULTADOS

---

<b>Figura 13.</b> Estudio <i>in vitro</i> de <i>splicing</i> de la variante de <i>BRCA1</i> c.75C>T (p.Pro25=). .....	92
<b>Figura 14.</b> Análisis mediante RT-PCR/EC del <i>splicing</i> alternativo en el entorno de los exones 9 y 10 en una LCL portadora de la variante <i>BRCA1</i> c.[594-2A>C; 641A>G] .....	95
<b>Figura 15.</b> Contribución de $\Delta 9,10$ y $\nabla 9,10$ (FL) a la expresión global de <i>BRCA1</i> determinada por PCR digital en un portador de la variante <i>BRCA1</i> c.[594-2A>C; 641A>G] (Carrier 1) y muestras control .....	96
<b>Figura 16.</b> Estudio <i>in vitro</i> de <i>splicing</i> de la variante de <i>BRCA1</i> c.5152+4T>A.....	98
<b>Figura 17.</b> Ejemplo de anotación manual de los datos de RNAseq.....	102
<b>Figura 18.</b> Ejemplo representativo de experimento RT-PCR/EC realizado entre los exones 1 y 3 de <i>PALB2</i> .....	102
<b>Figura 19.</b> La combinación de datos de RNAseq y RT-PCR/EC resulta fundamental para definir nuevos exones .....	103

<b>Figura 20.</b> Esquema seguido para la caracterización de eventos de <i>splicing</i> alternativo en <i>RAD51C</i> , <i>RAD51D</i> y <i>PALB2</i> .....	104
<b>Figura 21.</b> Ejemplo representativo de ensayo <i>RAD51C</i> E1 (int) E2 (int) analizado en una muestra de LCL.....	110
<b>Figura 22.</b> Ejemplo representativo de ensayo <i>RAD51C</i> E2E4 analizado en una LCL.....	110
<b>Figura 23.</b> Ejemplos representativos de ensayo <i>RAD51C</i> E4E5 en sangre, mama y ovario..	111
<b>Figura 24.</b> Caracterización de un evento de <i>splicing</i> alternativo no canónico. ....	112
<b>Figura 25.</b> Ejemplo representativo de ensayo <i>RAD51C</i> E7E9 realizado en una muestra de sangre.....	112
<b>Figura 26.</b> El <i>locus RAD51D</i> contiene dos exones 3 mutuamente excluyentes .....	112
<b>Figura 27.</b> Análisis por RNAseq del exón 3 de <i>RAD51D</i> .....	113
<b>Figura 28.</b> Ejemplo representativo de ensayo <i>RAD51D</i> E1E4 realizado en una LCL .....	114
<b>Figura 29.</b> Diagrama de cajas que muestra la contribución relativa de los exones E3 <sub>179</sub> y E3 <sub>119</sub> a la expresión global de <i>RAD51D</i> (E9E10) por dPCR .....	115
<b>Figura 30.</b> Ejemplo representativo de ensayo <i>RAD51D</i> E1E7 realizado en una muestra LCL. ....	120
<b>Figura 31.</b> Ejemplo representativo de ensayo E2E7 <i>RAD51D</i> realizado en ovario .....	120
<b>Figura 32.</b> Ejemplo representativo de ensayo <i>RAD51D</i> E2E8 realizado en ovario .....	121
<b>Figura 33.</b> Ejemplo representativo de ensayo <i>RAD51D</i> E3E5 realizado en una LCL .....	121
<b>Figura 34.</b> Ejemplo representativo de ensayo <i>RAD51D</i> E4E7 realizado en mama .....	121
<b>Figura 35.</b> Ejemplo representativo de ensayo <i>RAD51D</i> E6E7 realizados en ovario, una LCL y mama.....	122
<b>Figura 36.</b> Ejemplo representativo de ensayo <i>RAD51D</i> E6E9 realizado en una muestra de sangre.....	122
<b>Figura 37.</b> Análisis por RNAseq de <i>RAD51D</i> Δ3_5.....	123
<b>Figura 38.</b> Diagrama de Cajas que muestra los niveles de expresión de Δ3_5 de <i>RAD51D</i> (FAM) respecto la combinación del evento alternativo y el mensajero de referencia E3E4 (FAM + VIC) cuantificados por PCR digital.....	124
<b>Figura 39.</b> Ejemplos representativos de ensayo <i>PALB2</i> E1E3 realizados en mama, sangres, LCLs y ovario.....	128
<b>Figura 40.</b> Ejemplo representativo de ensayos <i>PALB2</i> E1E6 y E3E6 .....	129
<b>Figura 41.</b> Análisis por RNAseq de <i>PALB2</i> Δ4,5.....	129
<b>Figura 42.</b> Diagrama de Cajas que muestra los niveles de <i>PALB2</i> Δ4,5 respecto a la combinación del mensajero de referencia y al evento de <i>splicing</i> alternativo medidos por PCR digital .....	130
<b>Figura 43.</b> Ejemplo representativo de ensayo <i>PALB2</i> E5E8 realizado en una muestra LCL . .....	130

<b>Figura 44.</b> Ejemplo representativo de RT-PCR <i>PALB2</i> E6E8 analizados mediante electroforesis capilar en una muestra de LCL, tejido de mama y ovario .....	131
<b>Figura 45.</b> Ejemplo representativo de RT-PCR <i>PALB2</i> E7E10 analizados mediante electroforesis capilar en una muestra de LCL, tejido de mama y ovario .....	131
<b>Figura 46.</b> Ejemplo representativo de RT-PCR <i>PALB2</i> E9E12 analizados mediante electroforesis capilar en el tejido de mama.....	132
<b>Figura 47.</b> Ejemplo representativo de RT-PCR <i>PALB2</i> E10E13 analizados mediante electroforesis capilar en una muestra de LCL .....	132
<b>Figura 48.</b> Análisis por RNAseq de <i>PALB2</i> $\Delta 11$ , $\Delta 12$ y $\Delta 11,12$ .....	132
<b>Figura 49.</b> Diagrama de Cajas que muestra los niveles de <i>PALB2</i> E11E12 respecto al mensajero de referencia medidos por PCR digital.....	133

## **FIGURAS DISCUSIÓN**

---

<b>Figura 50.</b> Ejemplo representativo de RT-PCR <i>PALB2</i> E6E8 analizados mediante electroforesis capilar en una muestra de LCL, tejido de mama y ovario .....	146
--	-----

## **FIGURAS ANEXOS**

---

<b>Figura 51.</b> Transcritos del locus <i>RAD51C</i> descritos en ENSEMBL.....	165
<b>Figura 52.</b> Transcritos del locus <i>RAD51D</i> descritos en ENSEMBL .....	166
<b>Figura 53.</b> Transcritos del locus <i>PALB2</i> descritos en ENSEMBL.....	166
<b>Figura 54.</b> Diseño de los ensayos TaqMan para cuantificar la isoforma $\Delta 9.10$ de <i>BRCA1</i> ...	170
<b>Figura 55.</b> Diseño de los ensayos TaqMan para cuantificar la expresión de los exones 9 y 10 de <i>BRCA1</i> .....	170
<b>Figura 56.</b> Diseño de los ensayos TaqMan para cuantificar la expresión del evento $\Delta 3$ con respecto a la expresión global del locus <i>BRCA2</i> .....	171
<b>Figura 57.</b> Diseño de los ensayos TaqMan para cuantificar la expresión del evento $\Delta 3$ con respecto a la unión exón 3 y exón 4 del locus <i>BRCA2</i> .....	171
<b>Figura 58.</b> Diseño de los ensayos TaqMan para cuantificar la expresión del exón 3 <sub>119</sub> del locus <i>RAD51D</i> .....	172
<b>Figura 59.</b> Diseño de los ensayos TaqMan para cuantificar la expresión del exón 3 <sub>179</sub> del locus <i>RAD51D</i> .....	172
<b>Figura 60.</b> Diseño de los ensayos TaqMan para cuantificar la expresión de la unión de los exones 9 y 10 del locus <i>RAD51D</i> .....	172
<b>Figura 61.</b> Diseño de los ensayos TaqMan para cuantificar la expresión del evento $\Delta 3\_5$ del locus <i>RAD51D</i> .....	173
<b>Figura 62.</b> Diseño de los ensayos TaqMan para cuantificar la expresión de la unión de los exones 3 y 4 del locus <i>RAD51D</i> .....	173



<b>Figura 63.</b> Diseño de los ensayos TaqMan para cuantificar la expresión del evento $\Delta 4,5$ respecto a la expresión de la unión de los exones 4 y 5 del locus <i>PALB2</i> .....	174
<b>Figura 64.</b> Diseño de los ensayos TaqMan para cuantificar la expresión del evento $\Delta 11,12$ respecto a la expresión de la unión de los exones 4 y 5 del locus <i>PALB2</i> .....	174
<b>Figura 65.</b> Análisis del splicing de la variante <i>BRCA1</i> c.230C>T (p.Thr77Met).....	175
<b>Figura 66.</b> Análisis del splicing de las variantes <i>BRCA1</i> c.528G>A (p.Thr176=), c.570C>T (p.Thr190=), c.591C>T (p.Cys197=), c.773C>F (p.Pro258Arg), c.946A>G (p.Ser316Gly) y c.1242C>T (p.Asp414=) .....	176
<b>Figura 67.</b> Contribución relativa de $\Delta 9, 10$ a la expresión global de <i>BRCA1</i> determinada por PCR digital en portadores de las variantes <i>BRCA1</i> c.528G>A, c.570C>T, c.591C>T, c.773C>G, c.946A>G y c.1242C>T .....	177
<b>Figura 68.</b> Análisis del splicing de la variante <i>BRCA1</i> c.5175A>G (p.Glu1725=).....	178
<b>Figura 69.</b> Análisis del splicing de la variante <i>BRCA1</i> c.5333-2A>G (IVS21-2A>G) .....	179
<b>Figura 70.</b> Análisis del splicing de la variante <i>BRCA2</i> c.68-7T>A (IVS2-7T>A) mediante RT-PCR/EC .....	180
<b>Figura 71.</b> Contribución relativa de $\Delta 3$ a la expresión global de <i>BRCA2</i> determinada por PCR digital .....	181
<b>Figura 72.</b> Análisis del splicing de la variante <i>BRCA2</i> c.223G>C (p.Ala75Pro) mediante RT-PCR/EC .....	182
<b>Figura 73.</b> Contribución relativa de $\Delta 3$ a la expresión global de <i>BRCA2</i> determinada por PCR digital en portadores de las variantes <i>BRCA2</i> c.223G>C, controles negativos y un control positivo c.68-7T>A .....	182
<b>Figura 74.</b> Análisis del splicing de la variante <i>BRCA2</i> c.7007G>A (p.Arg2336His) mediante RT-PCR/Agarosa .....	183
<b>Figura 75.</b> Análisis del splicing de la variante <i>BRCA2</i> c.7007G>A (p.Arg2336His) mediante secuenciación Sanger .....	184
<b>Figura 76.</b> Análisis del splicing de la variante <i>BRCA2</i> c.7805+17G>A mediante RT-PCR /Agarosa.....	185
<b>Figura 77.</b> Análisis del splicing de la variante <i>BRCA2</i> c.7805+17G>A mediante Secuenciación Sanger .....	185
<b>Figura 78.</b> Análisis del splicing de la variante <i>BRCA2</i> c.8488-1G>A (IVS19-1G>A) .....	186
<b>Figura 79.</b> Análisis del splicing de la variante <i>BRCA2</i> c.8592C>T (p.Ala2864=) .....	187

# ÍNDICE DE TABLAS

---

## TABLAS DE INTRODUCCIÓN

---

<b>Tabla 1.</b> Algunas de las estimaciones de riesgo publicadas para <i>BRCA1</i> , <i>BRCA2</i> , <i>PALB2</i> , <i>RAD51C</i> y <i>RAD51D</i> .....	56
<b>Tabla 2.</b> Clasificación de las variantes genéticas propuesta por la agencia internacional para la investigación (IARC).....	59
<b>Tabla 3.</b> Algunos ejemplos representativos de reglas cualitativas incluidas en el algoritmo ENIGMA de clasificación clínica de variantes genéticas en <i>BRCA1</i> y <i>BRCA2</i> .....	61
<b>Tabla 4.</b> Algunos ejemplos de reglas cualitativas incluidas en el algoritmo ENIGMA de clasificación clínica de variantes genéticas en <i>BRCA1</i> y <i>BRCA2</i> .....	70

## TABLAS DE MATERIALES Y MÉTODOS

---

<b>Tabla 5.</b> Mezcla de la reacción RT-PCR .....	78
<b>Tabla 6.</b> Mensajeros de referencia para cada gen estudiado .....	79
<b>Tabla 7.</b> Ensayos TaqMan empleados en la cuantificación por PCR digital .....	84

## TABLAS DE RESULTADOS

---

<b>Tabla 8.</b> Estudios de <i>splicing</i> en portadores de variantes genéticas en <i>BRCA1</i> .....	99
<b>Tabla 9.</b> Estudios de <i>Splicing</i> en portadores de variantes genéticas en <i>BRCA2</i> .....	100
<b>Tabla 10.</b> Eventos de <i>splicing</i> alternativo en <i>RAD51C</i> Confirmados .....	105
<b>Tabla 11.</b> Eventos de <i>splicing</i> alternativo en <i>RAD51C</i> no Confirmados.....	108
<b>Tabla 12.</b> Detección de eventos de <i>splicing</i> alternativo en <i>RAD51C</i> (Confirmados y no Confirmados) estratificados por tipo de muestra biológica .....	109
<b>Tabla 13.</b> Eventos de <i>splicing</i> alternativo en <i>RAD51D</i> Confirmados .....	116
<b>Tabla 14.</b> Eventos de <i>splicing</i> alternativo en <i>RAD51D</i> no Confirmados .....	118
<b>Tabla 15.</b> Detección de eventos de <i>splicing</i> alternativo en <i>RAD51D</i> (Confirmados y no Confirmados) estratificados por tipo de muestra biológica .....	119
<b>Tabla 16.</b> Eventos de <i>splicing</i> alternativo en <i>PALB2</i> confirmados .....	125
<b>Tabla 17.</b> Eventos de <i>splicing</i> alternativo en <i>PALB2</i> no Confirmados.....	126
<b>Tabla 18.</b> Detección de eventos de <i>splicing</i> alternativo en <i>PALB2</i> (Confirmados y no Confirmados) estratificados por tipo de muestra biológica .....	127

## **TABLAS DE DISCUSIÓN**

---

<b>Tabla 19.</b> Implicaciones del splicing alternative en la clasificación clínica inicial de variantes genéticas en <i>RAD51C</i> , <i>RAD51D</i> y <i>PALB2</i> .....	154
--	-----

<b>Tabla 20.</b> Implicaciones del <i>splicing</i> en el diseño de estudios <i>in vitro/ex vivo</i> del <i>splicing</i> en <i>RAD51C</i> .....	156
--	-----

## **TABLAS DE ANEXOS**

---

<b>Tabla 21.</b> Cebadores utilizados en los estudios de splicing por electroforesis capilar y Secuenciación Sanger del locus <i>BRCA1</i> .....	167
--	-----

<b>Tabla 22.</b> Cebadores utilizados en los estudios de splicing por electroforesis capilar y Secuenciación Sanger del locus <i>BRCA2</i> .....	168
--	-----

<b>Tabla 23.</b> Cebadores utilizados en los estudios de splicing por electroforesis capilar y Secuenciación Sanger del locus <i>RAD51C</i> .....	168
---	-----

<b>Tabla 24.</b> Cebadores utilizados en los estudios de splicing por electroforesis capilar y Secuenciación Sanger del locus <i>RAD51D</i> .....	169
---	-----

<b>Tabla 25.</b> Cebadores utilizados en los estudios de splicing por electroforesis capilar y Secuenciación Sanger del locus <i>PALB2</i> .....	169
--	-----

# Resumen



## RESUMEN

---

Durante casi 20 años, *BRCA1* y *BRCA2* han sido los únicos genes que han proporcionado información útil para el asesoramiento de las familias con sospecha de padecer un síndrome hereditario de cáncer de mama y ovario (HBOC), a pesar de que los defectos germinales en estos genes no explican más de un 40% de estos síndromes.

La detección de una mutación patogénica en estos dos genes ha permitido individualizar el asesoramiento genético y proponer estrategias reductoras de riesgo más agresivas, como las cirugías profilácticas. Sin embargo, no siempre el resultado de los test genéticos es de carácter informativo, sino que en muchas ocasiones nos encontramos con mutaciones de significado clínico desconocido. Estas variantes son un problema importante, ya que no se puede concluir la relevancia clínica de la variante, y por tanto resulta complicada su clasificación clínica.

Hoy en día está bien establecido que los defectos genéticos en *PALB2* (cáncer de mama), y *RAD51C* o *RAD51D* (cáncer de ovario), tienen una contribución significativa al exceso de riesgo familiar (aunque las estimaciones de riesgo asociadas a cada uno de estos genes son imprecisas), lo que justifica su inclusión junto a *BRCA1* y *BRCA2* en paneles multigénicos con finalidad diagnóstica. Por ello, esperamos que el número de variantes genéticas de significado clínicamente desconocido en estos genes aumente de forma drástica en un futuro próximo, causando un serio problema en el asesoramiento genético.

En la presente tesis doctoral defendemos que la incorporación de los estudios *in vitro* de *splicing* a la rutina diagnóstica pueden aportar información importante para la clasificación clínica de variantes genéticas no sólo demostrando que ciertas variantes son patogénicas al alterar el *splicing*, sino también identificando variantes (en particular variantes sinónimas e intrónicas) que muy probablemente no lo sean al demostrarse que no afectan al *splicing*.

Además, en esta tesis doctoral defendemos la relevancia clínica de estudiar potenciales alteraciones de *splicing* asociadas a variantes genéticas inequívocamente no patogénicas (clasificación multifactorial), como es el caso de *BRCA1* c.594-2A>C (IVS9-2A>C) y *BRCA2* c.68-7T>A (IVS2-7T>A).

El estudio en *BRCA1* IVS9-2A>C (c.594-2A>C) ilustra perfectamente la relevancia del *splicing* alternativo a la hora de clasificar clínicamente una variante genética. El estudio *in vitro* de *splicing* alternativo realizado sobre esta variante ha permitido: 1) demostrar que no todas

las variantes genéticas localizadas en sitios consenso de *splicing* deben ser consideradas patogénicas en ausencia de estimaciones de riesgo directas y/o estudios experimentales, 2) demostrar que un codón de parada prematuro en *BRCA1* no es necesariamente patogénico, 3) demostrar que niveles normales de expresión de *BRCA1*  $\Delta 9,10$  confieren haplosuficiencia y, 4) demostrar que los estudios *in vitro* de *splicing* deben diseñarse teniendo en cuenta la existencia de *splicing* alternativo.

El estudio en *BRCA2* c.68-7T>A permite concluir que niveles de *BRCA2*  $\Delta 3$  hasta cuatro veces superiores a los observados en controles no se asocian con un aumento del riesgo de desarrollar cáncer.

Durante el desarrollo de esta tesis doctoral se ha tratado de demostrar que la caracterización detallada del *splicing* alternativo de un gen aporta información esencial para la correcta interpretación clínica de las variantes genéticas que se identifican en dicho gen; además de ser clave para el correcto diseño de los estudios de *splicing*. Por ello, consideramos esencial llevar a cabo una caracterización detallada del *splicing* alternativo en los genes candidato de susceptibilidad antes de que se incorporen a la clínica. De esta manera, consideramos que la caracterización del *splicing* alternativo en los genes de susceptibilidad *RAD51C*, *RAD51D* y *PALB2* llevado a cabo en este trabajo de tesis doctoral es una herramienta de enorme utilidad para la clasificación clínica de variantes genéticas en estos genes. En concreto, esta caracterización nos ha permitido: 1) identificar eventos de *splicing* alternativo que podrían tener impacto en la clasificación de variantes en zonas consenso de *splicing*, 2) identificar eventos de *splicing* alternativo que podrían tener un impacto sobre la clasificación de variantes que introducen codones de parada prematuros e 3) identificar eventos de *splicing* alternativo específicos de tejido.

# SUMMARY

---

For almost 20 years, only *BRCA1* and *BRCA2* genes have provided useful information in the genetic counseling of families with a suspected hereditary Breast and Ovarian cancer syndrome (HBOC), although germ-defects in these genes do not explain more than 40% of these syndromes.

The detection of a pathogenic mutation in these genes, allows the individualization of the genetic counseling and suggests more aggressive reducing-risk strategies, such as prophylactic surgeries. However, the result of the genetic test is not always informative, because often variants of unknown clinical significance are found. These variants represent an important problem, since the clinical relevance of these variants cannot be concluded; and therefore, their clinical classification is complicated.

Nowadays, it is well established that genetic defects in *PALB2* (Breast Cancer), and *RAD51C* or *RAD51D* (ovarian cancer), have a significant contribution to the familial risk excess (although the estimation risks for each gene are imprecise), which justifies their inclusion together with *BRCA1* and *BRCA2* genes in a multigenic panel for diagnostic. Therefore, we expect a dramatic increase in the number of unclassified variants in these genes, causing a serious problem in the genetic counseling.

In the present thesis project, we argue that the incorporation of *in vitro* splicing studies into the diagnostic routine may provide important information in the classification of genetic variants, not only demonstrating that certain pathogenic mutations altered the splicing, but also identifying variants (particularly synonymous and intronic variants) that probable are not pathogenic because they do not affect the splicing.

Also, in this thesis we defend the clinical relevance of studying potential splicing alterations associated with non-pathogenic variants (multifactorial classification), such as *BRCA1* c.594-2A>C (IVS9-2A>C) and *BRCA2* c.68-7T>A (IVS2-7T>A).

The *BRCA1* IVS9-2A>C (c.594-2A>C) study, perfectly illustrates the relevance of the alternative splicing in the clinical classification of the genetic variants. The *in vitro* alternative splicing study performed on this variant has allowed to: 1) demonstrate that not all genetic variants located in consensus splicing sites should be considered pathogenic in the absense of direct risk estimations and/or experimental studies, 2) demonstrate that a premature stop codon in *BRCA1* is not necessarily pathogenic, 3) demonstrate that normal expression levels of



*BRCA1* Δ9, 10 confer haplosufficiency and, 4) demonstrate that *in vitro* splicing studies should be designed, taking into account the existence of alternative splicing.

The *BRCA2* c.68-7T>A study allows to conclude that *BRCA2* Δ3 levels up to four times higher than observed in controls, are not associated with an increased of cancer risk.

During the development of this thesis, we tried to demonstrate that precisising the alternative splicing characterization of a gen, provides essential information for the correct clinical interpretation of the genetic variants found in this gen. Moreover, it is key in alternative splicing studies designed. Therefore, we consider essential to carry out a detailed characterization of the alternative splicing in the susceptible genes before are included in the clinic.

In this way, we consider that the characterization of alternative splicing in *RAD51C*, *RAD51D* and *PALB2* susceptible genes, carried out in this thesis project, is a very useful tool in the clinical classification of genetic variants found in these genes.

Specifically, this characterization has allowed us to: 1) identify alternative splicing events that may have an impact on variants classification located in the splicing consensus sites, 2) identify alternative splicing events that may have an impact on the classification of variants that introduce a premature stop codon, and 3) identify tissue-specific alternative splicing events.

# Glosario de abreviaturas



# GLOSARIO DE ABREVIATURAS

---

<b>ATM</b>	Ataxia telangiectasia mutated
<b>BRCA1</b>	Breast Cancer 1
<b>BRCA2</b>	Breast Cancer 2
<b>BRC</b>	Dominio de repetición en el exón 11 de <i>BRCA2</i>
<b>BRCT</b>	Dominio C-terminal de <i>BRCA1</i>
<b>BRIP1</b>	BRCA1 interacting protein C-terminal helicase 1
<b>CDH1</b>	Cadherin 1, type 1, E-cadherin
<b>cDNA</b>	Ácido desoxirribonucleico complementario
<b>CGE</b>	<i>Clonal Genome Evolution</i> (Evolución clonal del genoma)
<b>Chk2</b>	Checkpoint kinase 2
<b>CM</b>	Cáncer de mama
<b>CO</b>	Cáncer de ovario
<b>dbSNP</b>	Single Nucleotide Polymorphism Database
<b>DMC1</b>	<i>DNA meiotic recombinase 1</i> (Proteína de recombinación meiótica)
<b>DNA</b>	Ácido desoxirribonucleico
<b>dPCR</b>	<i>Digital PCR</i> (PCR digital)
<b>EC</b>	Electroforesis capilar
<b>ENIGMA</b>	Evidence based Network for the Interpretation of Germline Mutant Alleles
<b>ESE</b>	<i>Exonic Splicing Enhancer</i> (Potenciador del <i>splicing</i> exónico)
<b>ESS</b>	<i>Exonic Splicing Silencer</i> (Silenciador del <i>splicing</i> exónico)
<b>f</b>	Frecuencia poblacional
<b>FL</b>	<i>Full length</i> (transcrito de referencia)
<b>FAM</b>	<i>Fluorescein amidite</i>
<b>FS</b>	<i>Frameshift</i> (no mantiene la fase de lectura)
<b>GWAS</b>	<i>Genome-wide association study</i> (Estudios de asociación de genomas)
<b>HBOC</b>	<i>Hereditary breast and ovarian cancer</i> (cáncer de mama ovario hereditario)
<b>HCSC</b>	Hospital Clínico San Carlos
<b>HGVS</b>	Human Genome Variation Society
<b>HR</b>	Homologous recombination (Recombinación Homóloga)
<b>IARC</b>	<i>International Agency for Research on Cancer</i> (Agencia internacional de investigación en cáncer)
<b>ICLs</b>	Interstrand crosslinks
<b>ISE</b>	<i>Intronic splicing enhancer</i> (Secuencia intrónica potenciadora del <i>splicing</i> )
<b>ISS</b>	<i>Intronic splicing silencer</i> (Secuencia intrónica silenciadora del <i>splicing</i> )
<b>LCL</b>	Líneas celulares Linfoblastoides
<b>MAF</b>	Minor allele frequency (frecuencia mínima alélica)
<b>MMR</b>	Mismatch repair genes
<b>NGS</b>	<i>Next generation sequencing</i> (Secuenciación masiva de segunda generación)
<b>NMD</b>	Nonsense-mediated mRNA decay. Degradación del RNAm mediada por mutaciones terminadoras
<b>NLS</b>	<i>Nuclear Localization signal</i> (Señal de localización nuclear)
<b>No FS</b>	<i>No Frameshift</i> (Mantiene la fase de lectura)
<b>PALB2</b>	Partner and localizar of <i>BRCA2</i>
<b>PCR</b>	<i>Polymerase Chain Reaction</i> (Reacción en cadena de la Polimerasa)
<b>Pre-RNAm</b>	Ácido ribonucleico pre-mensajero

<b>PRS</b>	Polygenic risk score (Riesgo poligénico)
<b>PTC-NMD</b>	Premature translation-termination codon Nonsense-mediated mRNA decay
<b>RA</b>	Riesgo acumulado
<b>Rad50</b>	Proteína de reparación del DNA RAD50
<b>RAD51C</b>	Proteína de reparación del DNA RAD51C
<b>RAD51D</b>	Proteína de reparación del DNA RAD51D
<b>RNA</b>	Ácido ribonucleico
<b>RNA<sub>m</sub></b>	Ácido ribonucleico mensajero
<b>RNAseq</b>	RNA sequencing (Secuenciación RNA)
<b>RR</b>	Riesgo relativo
<b>RT-PCR</b>	Reverse transcription polymerase chain reaction
<b>SF</b>	Splicing fraction
<b>SNP</b>	Single Nucleotide Polymorphism
<b>Splicing</b>	Corte y empalme de exones
<b>STAR</b>	Ultrafast Universal RNA-seq aligner
<b>STK11</b>	Serine/threonine kinase 11
<b>TP53</b>	Tumor protein p53
<b>VIC</b>	Fluorescent dye
<b>VUS</b>	<i>variant of uncertain significance</i> (variante de significado clínico desconocido)
<b>Wt</b>	<i>Wild type</i> (Nativo)
<b>XRCC2</b>	Proteína de reparación del DNA XRCC2
<b>XRCC3</b>	Proteína de reparación del DNA XRCC3

# Introducción



# INTRODUCCIÓN

---

## 1. SUSCEPTIBILIDAD GENÉTICA AL CÁNCER DE MAMA Y OVARIO: *BRCA1*, *BRCA2*, *PALB2*, *RAD51C* Y *RAD51D* EN CONTEXTO.

### 1.1. El cáncer es una enfermedad genética. El modelo CGE

El reciente desarrollo de nuevas tecnologías de análisis genético, en particular la secuenciación masiva en paralelo (NGS, del inglés *Next-Generation Sequencing*), ha permitido la caracterización del genoma de numerosos tumores sólidos individuales con un nivel de resolución sin precedentes (Vogelstein B et al., 2013). En conjunto, estos análisis han demostrado de forma concluyente la hipótesis de que el cáncer es en último término un proceso evolutivo impulsado por la selección natural actuando sobre mutaciones somáticas preexistentes adquiridas de forma estocástica (Fox EJ et al., 2016).

Este modelo darwiniano de carcinogénesis (modelo CGE, del inglés *Clonal Genome Evolution*) predice que las alteraciones genéticas identificadas en las muestras tumorales pueden clasificarse en mutaciones *driver* y mutaciones *passenger*, en función de su papel evolutivo (Greenman C et al., 2007). Por definición, las mutaciones *driver* proporcionan (o han proporcionado en algún momento de la evolución tumoral) una ventaja adaptativa fundamental, mientras las mutaciones *passenger* no tienen ningún valor adaptativo, aunque puedan contener no obstante información clínicamente relevante, como puede ser el tipo de inestabilidad genética operando en el tumor (Alexandrov LB et al., 2013), o el potencial del tumor para evolucionar en el futuro adaptándose a nuevos cambios ambientales (como, por ejemplo, la aparición de poblaciones resistentes a determinados fármacos anticancerígenos, Vogelstein B et al., 2013).

Por supuesto, las mutaciones *driver* no ocurren con igual probabilidad en todos los genes del genoma. Por el contrario, se concentran en unos pocos genes que, al alterarse, permiten a las células evadir las estrictas reglas de organización interna que caracterizan a los organismos metazoos. Estimaciones recientes indican que el número de genes humanos, capaz de comportarse como *drivers* del proceso carcinogénico es relativamente bajo: probablemente menos de 200 genes de un genoma que contiene aproximadamente 20.000 genes codificadores de proteínas (Vogelstein B et al., 2013).



Las mutaciones *driver* deben permitir a las células tumorales (o en proceso de llegar a ser tumorales) adquirir habilidades tales como la capacidad de proliferación celular rápida, indefinida e independiente de señales externas (lo que a su vez implica alterar los mecanismos de señalización entre la membrana y el núcleo celular, suprimir los mecanismos de senescencia inducidos por estrés replicativo, reactivar mecanismos preservadores de la longitud telomérica, y modificar el metabolismo energético), la capacidad de evadir los controles del sistema inmune, la capacidad de suprimir los mecanismos inductores de apoptosis (mecanismos exógenos y endógenos), la capacidad de neo-vascularización del tejido tumoral en crecimiento (angiogénesis), y la capacidad de movilidad, extravasación, invasión y metástasis (Hanahan D et al., 2011). De acuerdo con el modelo CGE, la adquisición de todas estas características se alcanza mediante un proceso secuencial de expansiones clonales, cada una de ellas motivada por la adquisición de una mutación en un gen *driver* que confiere ciertas ventajas selectivas a un sub-clon de células. Probablemente, el número de mutaciones *drivers* (y por tanto el número de expansiones clonales) que “explican” cada tumor individual en el momento del diagnóstico es relativamente bajo. Algunos estudios lo cifran en torno a 10 mutaciones *driver* por tumor (Vogelstein B et al., 2013).

En cualquier caso, es fácil comprender que en un proceso de este tipo la probabilidad de que llegue a emerger un tumor clínicamente aumenta, si aumenta la tasa de mutaciones. De hecho, hoy en día está firmemente establecida la noción de que una de las características de las células tumorales es la adquisición temprana de inestabilidad genética (Hanahan D et al., 2011).

### 1.2. Susceptibilidad genética al Cáncer

El modelo CGE describe un proceso eminentemente estocástico. Sin embargo, de forma en apariencia paradójica, el cáncer es una enfermedad que tiende a ocurrir en individuos genéticamente predispuestos. Numerosos estudios epidemiológicos realizados desde mediados del siglo XX han venido indicando de forma consistente que los familiares en primer grado de afectos de cáncer tienen un riesgo de desarrollar la enfermedad 2 ó 3 veces superior a la población general, lo que parece implicar un componente de susceptibilidad genética al cáncer importante (no obstante, la agregación familiar también es compatible con factores no genéticos como los ambientales y/o los infecciosos).

En un trabajo seminal publicado en el año 2000, el grupo de Lichtenstein y colaboradores logró discriminar la importancia relativa de los factores ambientales y genéticos como causantes del cáncer (Lichtenstein P et al., 2000). Para ello, analizaron la incidencia de

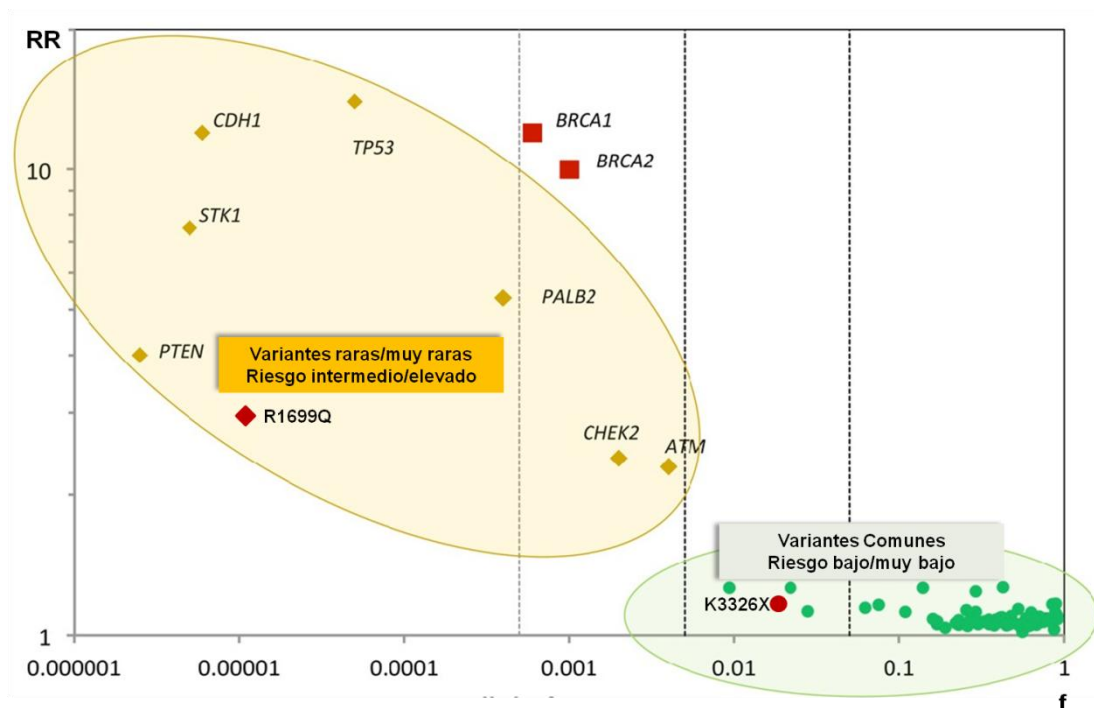
cáncer en 44.788 parejas de gemelos (daneses, suecos y finlandeses), comparando la concordancia observada en gemelos monocigóticos frente a la observada en gemelos dicigóticos. De este modo, los investigadores pudieron estimar la Heredabilidad de numerosos cánceres. La Heredabilidad es una medida de la carga genética asociada a un fenotipo dado. De forma rigurosa, la Heredabilidad se define como la proporción (normalmente expresada en tanto por uno) de la variabilidad fenotípica observada en una población (en este caso concreto desarrollar o no desarrollar cáncer) que puede atribuirse a una variabilidad genética en dicha población. Por tanto, una Heredabilidad de 0 equivale a ausencia de componente genético, mientras que una Heredabilidad de 1 equivale a un fenotipo completamente determinado genéticamente. De acuerdo con el estudio de Lichtenstein, la Heredabilidad de algunos de los cánceres más comunes en la población oscila entre 0.22 (cáncer de ovario) y 0.42 (cáncer de próstata), con valores de 0.26, 0.27, 0.35 y 0.36 para los cánceres de pulmón, mama, colorrectal, y páncreas respectivamente (Lichtenstein P et al., 2000).

### 1.3. Heredabilidad y arquitectura genética subyacente

Desde el punto de vista de la traslación a la clínica (asesoramiento genético en cáncer familiar), mucho más relevante que el valor absoluto de la Heredabilidad de un cáncer, es la arquitectura genética subyacente a dicha Heredabilidad. Por arquitectura genética entendemos el número total de *loci* asociados al riesgo de desarrollar la enfermedad, la frecuencia poblacional de los alelos de riesgo existentes en cada *loci*, la magnitud de la asociación al riesgo para cada uno de los alelos individuales, y la existencia o no de interacciones genéticas entre *loci*. Lógicamente, cuanto mayor sea la magnitud de la asociación con el riesgo atribuida a un *locus* individual, y mayor la frecuencia poblacional de los alelos de riesgo presentes en dicho *locus*, mayor será la contribución individual de dicho locus a la Heredabilidad (y más fácil su traslación a la clínica).

De forma un tanto arbitraria, (aunque justificable desde un punto de vista histórico y metodológico) la compleja arquitectura genética subyacente a la Heredabilidad del cáncer de mama y ovario (u otros cánceres) se suele descomponer en 3 componentes principales: genes donde se localizan variantes genéticas muy raras en la población pero asociadas a riesgo muy elevado de desarrollar la enfermedad (muchos de estos genes se identificaron en el periodo 1985-1995 mediante análisis de ligamiento y clonaje posicional), genes donde se localizan variantes genéticas relativamente raras de riesgo “intermedio” (aunque sería más correcto referirse a estimaciones de riesgo imprecisas), y genes donde se localizan variantes genéticas comunes en la población asociadas a incrementos muy modestos en el riesgo de desarrollar

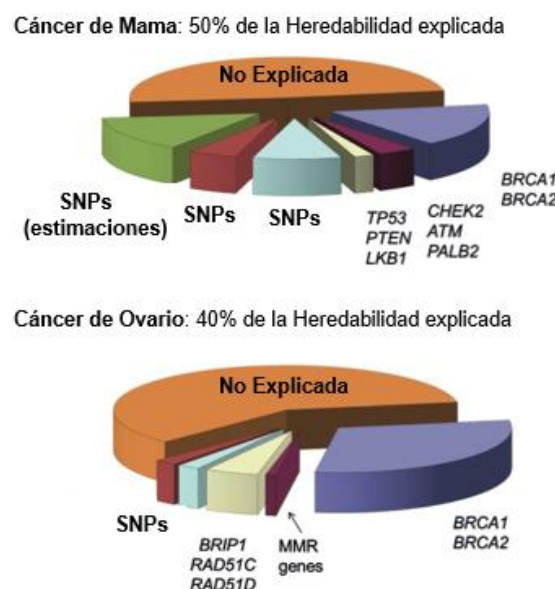
cáncer (la mayoría localizados en el periodo 2005-2015 mediante estudios GWAS, del inglés *genome-wide association studies*). En nuestra opinión, en el momento actual tiene más lógica agrupar los genes de riesgo “elevado” e “intermedio” en un único componente de la Heredabilidad definido por genes donde se localizan variantes raras o muy raras de riesgo intermedio/elevado, y así lo reflejamos en la Figura 1. De este modo, está firmemente establecido que *BRCA1*, *BRCA2* y *PALB2* contribuyen a la Heredabilidad del cáncer de mama Figura 1, mientras que *BRCA1*, *BRCA2*, *RAD51C* y *RAD51D* contribuyen a la Heredabilidad del cáncer de Ovario (Figura 2).



**Figura 1.** La figura representa los distintos *loci* de susceptibilidad al cáncer de mama en función de su frecuencia poblacional ( $f$ ) y su asociación con el riesgo relativo ( $RR$ ). Se puede observar una correlación inversa casi perfecta entre  $RR$  y  $f$  que se justifica por motivos tanto epidemiológicos (variantes comunes de riesgo elevado serían incompatibles con la incidencia de cáncer en la población general) como metodológicos (con los métodos analíticos actuales es muy difícil llegar a detectar variantes raras asociadas a riesgo bajo). *BRCA1* y *BRCA2* parecen desviarse de la norma de la correlación inversa, con unos  $RR$  muy elevados en relación con la  $f$  en la población. Esta observación refleja la posición especialmente preponderante que *BRCA1* y *BRCA2* ocupan en la susceptibilidad genética al cáncer de mama. *TP53*, *CDH1*, *STK11* y *PTEN* se asocian con riesgos muy elevados de cáncer de mama, pero en el contexto de síndromes hereditarios muy característicos que nada tienen que ver clínicamente con el síndrome mama/ovario asociado a las mutaciones en *BRCA1* y *BRCA2*. En el caso de los *loci* de riesgo intermedio/elevado, la  $f$  se refiere a la frecuencia de variantes genéticas causales en los genes indicados. En el caso de los *loci* de riesgo bajo/muy bajo, la  $f$  se refiere a la frecuencia del SNP asociado al riesgo en estudios GWAS, que no tiene porqué ser la variante genética causal. La figura también trata de representar otro aspecto relevante de la arquitectura genética del cáncer. El riesgo viene determinado por variantes genéticas concretas, de tal modo que aún en genes denominados de alta penetrancia como *BRCA1* y *BRCA2*, puedan existir variantes genéticas de riesgo intermedio (*BRCA1* R1699Q) o bajo (*BRCA2* K3326X). La figura no representa adecuadamente el error asociado a las

estimaciones del riesgo. En algunos casos (por ejemplo *PALB2*, *CHEK2*, y *ATM*) el error puede ser muy elevado. (Figura adaptada de Ghoussaini et al, 2013).

La identificación de genes de riesgo “intermedio/elevado”, y la caracterización precisa de su contribución a la Heredabilidad es una de las áreas de la investigación en genética del cáncer de mama y ovario más activa en la actualidad. La convergencia de numerosos estudios genéticos y funcionales han permitido: 1) establecer el papel central de *BRCA1* y *BRCA2* en los procesos de reparación del DNA por recombinación homóloga (Roy et al, 2012); 2) caracterizar una ruta de detección y reparación de enlaces covalentes entre hebras complementarias de DNA (ICLs, del inglés inter-strandcrosslink) dependiente de recombinación homóloga denominada ruta Anemia de Fanconi-BRCA (Kim et al, 2012); 3) reconocimiento de una base genética común para la Anemia de Fanconi (FA) y la susceptibilidad genética al cáncer de mama y ovario (D'Andrea AD. 2010); y 4) establecer el principio según el cual cualquier gen implicado en reparación de ICLs por recombinación homóloga (genes FA/BRCA/HR) debe ser evaluado como posible gen de susceptibilidad al cáncer de mama y ovario (Levy-Lahad et al, 2010). Este marco teórico ha permitido proponer (con mayor o menor grado de evidencia) la existencia de numerosos genes de riesgo intermedio/alto implicados en susceptibilidad genética al cáncer de mama y/u ovario (Walsh et al, 2011; Kiiski et al, 2014; Kurian et al, 2014; Tung et al, 2016), siendo los ya comentados *BRCA1*, *BRCA2*, *RAD51C*, *RAD51D*, *CHEK2*, *ATM* y *BRIP1* los que cuentan hoy en día con mayor grado de evidencia (Easton et al, 2015). (Figura 2).



**Figura 2.** Los gráficos representan para los casos particulares del cáncer de mama y el cáncer de ovario, el porcentaje de la Heredabilidad que se considera explicada por los *loci* de susceptibilidad conocidos en la actualidad, diferenciando la contribución de las variantes de alto riesgo, riesgo intermedio, y bajo riesgo. En el caso del cáncer de mama, la contribución de *BRCA1* y *BRCA2* es muy importante (~20%), pero la contribución colectiva de los SNPs

conocidos es semejante ( $\approx 16\%$ ), y probablemente superior, si tenemos en cuenta las estimaciones GWAS en relación a los SNPs pendientes de identificación (señalados en verde). En el caso del cáncer de ovario la contribución de *BRCA1* y *BRCA2* es incluso mayor ( $\approx 40\%$ ). Por contra, la contribución colectiva de los SNPs es muy pequeña. Probablemente, esto último sólo refleje el hecho de que los estudios GWAS realizados hasta la fecha en cáncer de ovario no han tenido suficiente potencia estadística. Se espera que los genes de riesgo intermedio puedan llegar a explicar gran parte de la Heredabilidad no explicada en la actualidad, pero lo cierto es que con los conocimientos actuales su contribución es muy modesta, con la posible excepción del cáncer de ovario. (Figura adaptada de Bahcall. OC. Nat Genet. 2013).

Aunque asociemos frecuentemente genes con niveles de riesgo, lo cierto es que el riesgo es conferido por variantes genéticas concretas, y diversas variantes en un mismo gen pueden asociarse con niveles de riesgo muy diversos. Así, por ejemplo, hoy en día está bien establecido que distintas mutaciones en *BRCA1* y *BRCA2* se pueden asociar con niveles de riesgo significativamente distintos, tanto en cáncer de mama como en cáncer de ovario (Rebbeck TR et al., 2015). Se han identificado variantes genéticas en *BRCA1* asociadas a RR en torno a 3 (propio de los denominados genes de riesgo intermedio) (Spurdle AB et al., 2012); y variantes genéticas en *BRCA2* asociadas a RR menores de 1.5 (propio de las denominadas variantes de bajo riesgo) (Meeks HD et al., 2016) (Figura 1). Del mismo modo, es previsible que distintas variantes genéticas en otros genes como *PALB2*, *RAD51C* o *RAD51D* se asocien con distintos niveles de riesgo.

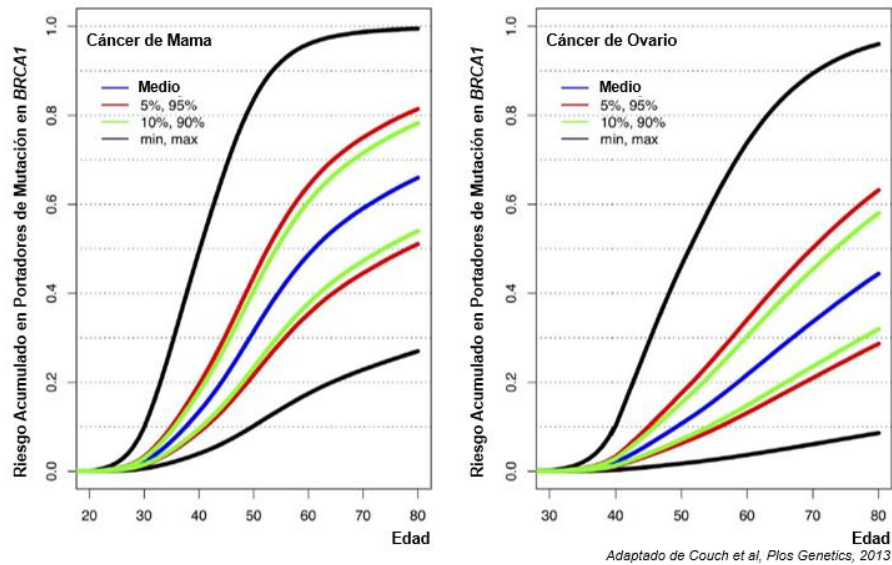
### 1.4. El Modelo Poligénico de Susceptibilidad al Cáncer de Mama y Ovario

En esencia, el modelo poligénico establece que para determinar con precisión el nivel de susceptibilidad genética al cáncer de mama y ovario a un individuo (y por tanto su manejo clínico) se deberían analizar todos los *loci* implicados (independientemente del RR conferido por cada uno de ellos). Si esto fuese posible, podríamos en principio estratificar toda la población en función de su riesgo genético de un modo mucho más preciso que el actual y, muy importante, la estratificación sería independiente de los antecedentes familiares.

Por supuesto, antes de trasladar el modelo poligénico a la clínica debemos conocer en detalle la arquitectura genética subyacente, lo que implica no sólo identificar todos los *loci* implicados y estimar los correspondientes riesgos, sino establecer las posibles interacciones entre ellos. En la actualidad existen evidencias parciales que sugieren que el modelo poligénico en cáncer de mama y ovario es válido y puede llegar a ser útil en la clínica.

Por ejemplo, se ha estimado que el cálculo de un RR poligénico (PRS, del inglés polygenic risk score) resultado de multiplicar los RR de 90 SNPs permite estratificar la población en niveles de riesgo de cáncer de mama que pueden tener relevancia clínica. En base al PRS, el 5% de las mujeres tienen un RR de 2 de desarrollar la enfermedad, y el 0.7% alcanzan un RR de 3 (Mavaddat N et al., 2015). El PRS es capaz de estratificar el riesgo tanto en mujeres sin historia familiar como en mujeres con historia familiar (Mavaddat N et al., 2015). Se han definido PRSs con capacidad de estratificar el riesgo en cáncer de ovario (Jervis S et al., 2014).

Muy importante, el modelo es válido cuando se introducen genes de alta penetrancia. En concreto, en cáncer de mama se puede combinar multiplicativamente un PRS basado en SNPs con el RR conferido por mutaciones patogénicas en *BRCA1* y *BRCA2*. En este caso, es común referirse a los SNPs como “modificadores del riesgo” en portadores de mutación. Las evidencias actuales indican que, una vez estratificados los tumores en RE positivos y negativos, el efecto modificador del riesgo asociado al PRS es equivalente en portadoras de mutación patogénica en *BRCA1*, en *BRCA2*, y el observado en la población general (Kurian AW et al., 2015). Lógicamente, el efecto modificador del PRS expresado en términos de riesgo absoluto es mucho mayor en portadores. El PRS permite una estratificación del riesgo en portadores de mutación patogénica en *BRCA1* y *BRCA2* que puede tener implicaciones clínicas. Por ejemplo, utilizando un PRS del cáncer de mama basado únicamente en 10 SNPs, se puede estratificar el riesgo absoluto de las portadoras de mutación en *BRCA1* (Figura 3) de tal modo que el grupo de mayor riesgo alcanza un riesgo acumulado de desarrollar cáncer de mama del 100% (penetrancia completa), mientras que el grupo de menor riesgo tiene un riesgo acumulado de alrededor del 25% (Couch et al, 2013). Del mismo modo, utilizando un PRS de cáncer de ovario basado exclusivamente en 7 SNPs (Figura 3), se puede estratificar el riesgo de tal modo que el grupo de mayor riesgo tiene un riesgo acumulado de más del 90%, mientras el grupo de menor riesgo no llega al 10% (Couch FJ et al., 2013).



**Figura 3.** Las gráficas representan el riesgo acumulado de desarrollar cáncer de mama y cáncer de ovario a lo largo de la vida en mujeres portadoras de mutaciones patogénicas en *BRCA1*. En azul se representa la estimación de riesgo media (la que se utiliza en la actualidad para asesorar a todas las mujeres portadoras de mutación patogénica). Como se observa, la incorporación de un componente poligénico (en este caso limitado a 10 SNPs en el caso del cáncer de mama y 7 SNPs en el caso del cáncer de ovario) permite una estratificación del riesgo que puede tener implicaciones clínicas (Figura adaptada de Couch et al., Plos Genetics, 2013).

Probablemente, el PRS modificará de forma equivalente el riesgo conferido por mutaciones en otros genes de riesgo intermedio/alto, como *PALB2*, *RAD51C* y *RAD51D*, aunque hasta donde sabemos, este extremo no se ha comprobado formalmente.

Por supuesto, para desarrollar un modelo que permita establecer con precisión el RR de desarrollar cáncer, habrá que combinar el modelo poligénico en desarrollo con factores no genéticos (ambientales, hábitos de vida) que pueden ser tan o más determinantes que los factores genéticos. Los pocos datos existentes se restringen a la interacción de factores no genéticos con portadores de mutación en *BRCA1* y *BRCA2*, y no son concluyentes (Friebel TM et al., 2014).

## 2. *BRCA1*, *BRCA2*, *PALB2*, *RAD51C* Y *RAD51D* COOPERAN EN LA REGULACIÓN DE LA HOMEOSTASIS DEL DNA DE DOBLE HEBRA

### 2.1. *BRCA1*

El gen *BRCA1* (17q21) fue descubierto en 1994 por Mary-Claire King y colaboradores. El mensajero considerado de referencia (ENST00000357654.7) se compone de 23 exones, y contiene una secuencia codificante (repartida en 22 exones) que se traduce en una proteína de 1863 aminoácidos (~220Kda).

Esta proteína presenta dos dominios funcionales, un dominio llamado anillo de zinc (Ring) en el extremo amino-terminal, y dos dominios BRCT en el extremo Carboxi-terminal.

El dominio Ring del extremo amino-terminal presenta motivos de Histidina y Cisteína conservados que van a mediar la unión de *BRCA1* con su proteína homóloga *BARD1* que al igual que *BRCA1*, presenta la misma estructura: 2 dominios BRCT en el extremo C-terminal y un dominio de Ring en el extremo amino-terminal. La unión de *BRCA1*-*BARD1* será necesaria para formar un complejo E3 ubiquitina ligasa que mediará en la regulación de la reparación homóloga del DNA (Baer, R. and T. Ludwig, 2002; Ruffner, H., et al., 2001).

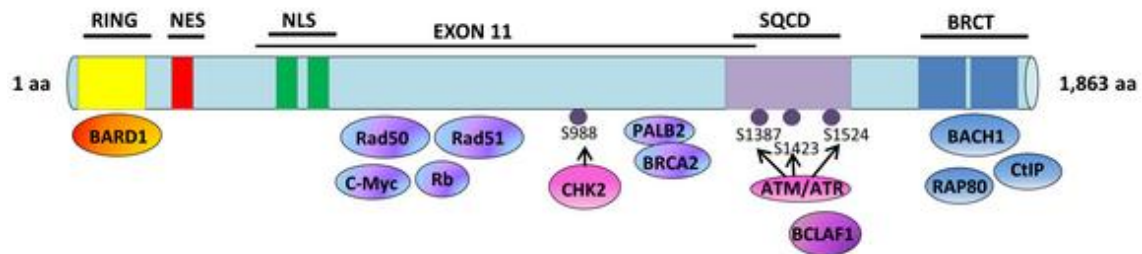
En el extremo C-terminal presenta dos dominios BRCT que también van a jugar un papel muy importante en la reparación homóloga (Wang, B., et al., 2007). Estos dominios BRCT reconocen y se unen a fosfopéptidos de otras proteínas que participan en la cascada de señalización en la reparación homóloga como *BACH1* y *CtIP* (Williams, R.S., et al., 2004).

La región que contiene los exones 11 y 13 comprende el 65% de las secuencias peptídicas de *BRCA1*. A diferencia de los dominios RING y BRCT, esta región no adopta ninguna estructura globular (es una región proteica intrínsecamente desordenada), pero contiene 2 señales de localización nuclear (NLS), un dominio SQCD y un dominio de exportación nuclear. Mutaciones en el dominio NLS impiden la interacción de *BRCA1* con la importina- $\alpha$ , lo que imposibilita la traslocación de *BRCA1* al núcleo que es necesario para la reparación del DNA. Además, muchas proteínas que participan en diversos procesos celulares como: las proteínas de reparación *Rad50* y *Rad51*, factores de transcripción como *c-Myc* y reguladores del ciclo celular; se unen en esta región que comprenden los exones 11-13 (Clark SL. et al., 2012).

En cuanto al dominio SQCD, también resulta imprescindible para la reparación del DNA, ya que contiene residuos de serina-glutamina (SQ) y threonina-glutamina (TQ) que son diana para la fosforilación por parte de 2 proteínas quinasas (*ATM* ó *ATR*) que inducirán la



fosforilación de *BRCA1* como respuesta ante un daño en la cadena de DNA (Traven, A. and J. Heierhorst, 2005). A continuación, se muestra un esquema de la estructura de la proteína *BRCA1* (Figura 4):



**Figura 4. Esquema de la proteína *BRCA1*.** Se identifican las distintas regiones que conforman la estructura de *BRCA1* y la asociación con otras proteínas que participan en la vía de reparación. *Katy S. Orr and Kieran I. Savage, 2015.*

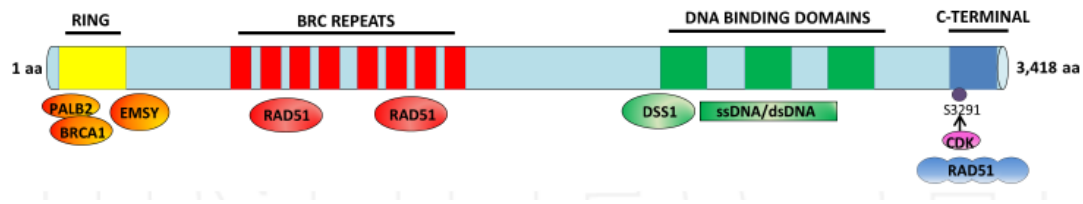
## 2.2. *BRCA2*

El gen *BRCA2* (13q12.3) fue descubierto por Wooster y colaboradores en 1995. El mensajero de referencia (ENST00000380152.7) se compone de 27 exones, y codifica una secuencia codificante (repartida en 27 exones) que se traduce en una proteína de 3418 aminoácidos.

La única homología que presenta con otros genes es un dominio con repeticiones BRC situado en el exón 11. Presenta 8 motivos de 30 residuos llamados BRC que codifica el exón 11, y que parecen estar muy conservados entre especies de mamíferos, lo que parece indicar que resulta esencial para su función (Bignell et al., 1997). De hecho, estudios de cristalografía han observado que los dominios BRC median la unión de *BRCA2* con otra proteína de reparación Rad51.

Además, *BRCA2* presenta un dominio de unión al DNA que le permiten asociarse con cadenas tanto de DNA sencillo como doble (Holloman, W.K et al., 2011; Jensen, R.B et al., 2010; Yang, H et al., 2002). En su extremo amino-terminal presenta una región que interacciona con la proteína PALB2, y además de jugar un papel importante en la reparación homóloga (debido a su unión con Rad51), también está involucrado en la activación transcripcional. En cuanto a su dominio C-terminal, puede unirse con múltiples moléculas de Rad51, a diferencia del dominio BRC del exón 11 que sólo se puede unir con una molécula de Rad51.

A continuación, se representa un esquema de la estructura de la proteína BRCA2 (Figura 5):

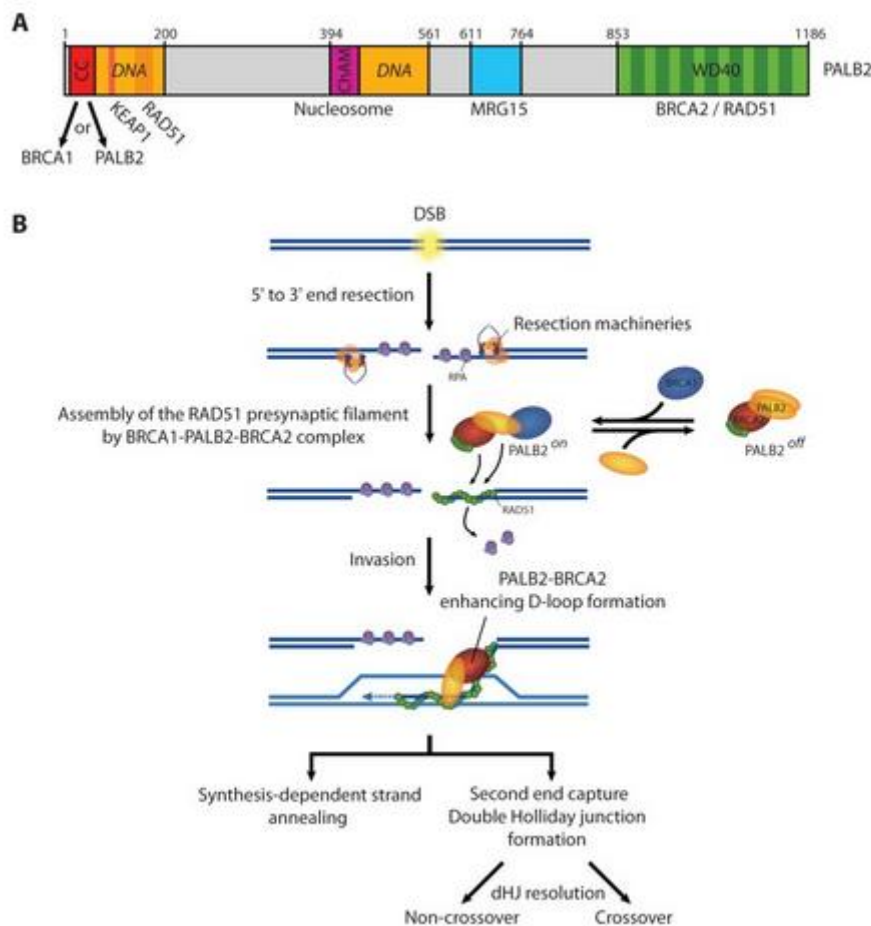


**Figura 5. Esquema de la proteína BRCA2.** Se identifican las distintas regiones que conforman la estructura de BRCA2 y la asociación con otras proteínas que participan en la vía de reparación. *Katy S. Orr and Kienan I. Savage, 2015*

## 2.3. PALB2

El gen *PALB2* se encuentra localizado en el brazo corto del cromosoma 16. El mensajero de referencia ENST00000261584.8 se compone de 13 exones codificantes que se traducen en una proteína de 1186 aminoácidos (131KDa), que participa en la supresión tumoral junto con la proteína de reparación BRCA2 (Figura 6). *PALB2* fue originalmente identificado como proteína de interacción con BRCA2 (Xia B et al., 2006; Sy SM et al., 2009). Esta función es necesaria para que *BRCA2* desempeñe su función. Además, se sabe que *PALB2* también interacciona con *BRCA1* (Zhang F et al., 2009).

*PALB2* se co-localiza con *BRCA2* en el núcleo fomentando su estabilidad, y facilita la unión de los genes *BRCA*, por lo que parece jugar un papel crítico en la ruta de reparación del DNA por recombinación homóloga (Xia B. et al., 2006). A continuación, se representa cómo *PALB2* interacciona con los genes *BRCA* formando un complejo (Figura 6). Aún se desconoce qué proteínas de las que interaccionan con el dominio BRCT de *BRCA1* (BRIP1/FANCI, CCDC98-RAP80, o CtIP) existen en el núcleo donde se forma el complejo BRCA1/PALB2/BRCA2. Además, *BRCA1* puede ser reclutado para la reparación mediante su interacción con el complejo MRN (MRE11/RAD50/NBS1), o mediante su interacción con el complejo CCDC98/RAP80. Aún se desconoce cuál de las dos interacciones es responsable del reclutamiento de PALB2/BRCA2/RAD51 (Tischkowitz M T y Xia B, 2010).



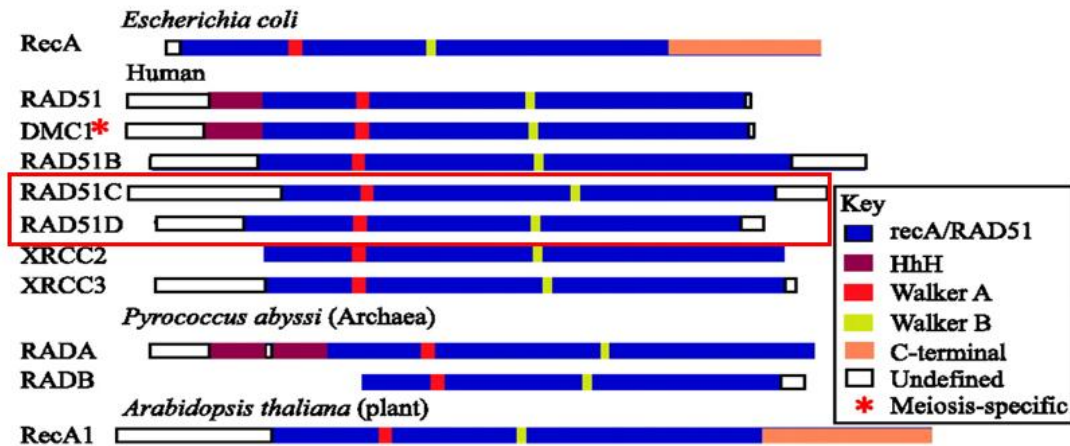
**Figura 6. Esquema de PALB2.** A. Estructura de PALB2 y sus dominios de interacción con otras proteínas. B. Modelo de interacción entre PALB2 y los genes BRCA para la reparación de roturas de doble hélice. (Figura adaptada de Buisson R and Masson JY, 2012).

Las mutaciones a nivel germinal del gen *PALB2* que provocan la pérdida de función en los dos alelos, tanto en el de la variante como en el normal, se asocian con la anemia de Fanconi tipo N. Esta enfermedad, suprime la función de la médula ósea y provoca la caída en los niveles de glóbulos rojos, leucocitos y plaquetas. Sin embargo, las mutaciones que afectan a un solo alelo (monoalélicas), se asocian con un incremento en el riesgo de padecer cáncer de páncreas y/o mama (Xia B et al., 2007; Reid S et al., 2007; Foulkes WD et al., 2007; Rahman N et al., 2007; Tischkowitz M y Xia B, 2010; Slater EP et al., 2010).

## 2.4. RAD51C y RAD51D

En humanos, la familia de genes *recA/RAD51* implicados en recombinación homóloga cuenta con 7 miembros (*RAD51*, *DMC1*, *RAD51C*, *XRCC3*, *RAD51B*, *RAD51D* y *XRCC2*), todos ellos parálogos originados a partir de un gen ancestral común por duplicación y divergencia

(Lin et al, 2006). Tal y como se indica en la figura 7, las proteínas RecA/RAD51 presentan un dominio central de unos 230 aminoácidos con alta homología (región RecA/RAD51 señalada en azul en la figura) y divergencia en las regiones N-terminal y C-terminal. La región central conservada incluye los dominios Walker A y B, esenciales para la actividad ATPasa de estas proteínas (Figura 7).



**Figura 7. Familia de genes *recA/RAD51*.** En rojo se destacan los genes *RAD51C* y *RAD51D*, que son miembros de la familia de genes *RecA/RAD51*, esenciales en todos los procesos fisiológicos de recombinación del DNA (figura adaptada de Lin et al, 2006).

Aunque todos los genes de la familia están funcionalmente relacionados con la recombinación homóloga, se piensa que han evolucionado para especializarse en diversas funciones. Por ejemplo, múltiples monómeros de la proteína codificada por el gen *RAD51* forman filamentos que protegen el DNA de hebra sencilla y catalizan su recombinación con secuencias homólogas, cumpliendo una función esencial tanto en reparación del DNA como en Meiosis. De forma similar, múltiples monómeros de *DMC1* interaccionan con DNA de hebra sencilla, pero la función de *DMC1* es específica de Meiosis (Lin et al, 2006).

Al parecer, los otros 5 miembros de la familia *RecA/RAD51* funcionan interaccionando entre sí para formar complejos multiproteicos que modulen la actividad de *RAD51* y *DMC1*, protegen el DNA de hebra sencilla presente en la horquilla de replicación, y remodelan los filamentos de *RAD51* y *DMC1* para estimular el proceso de recombinación (Somyajit et al, 2015; Taylor et al, 2015; Su et al, 2017). Se han descrito al menos dos de estos complejos: BCDX2 (formado por *RAD51B*, *RAD51C*, *RAD51D* y *XRCC2*) y CX3 (formado por *RAD51C* y *XRCC3*). El complejo BCDX2 parece actuar en fases iniciales del proceso de recombinación, cooperando con *BRCA2* en la formación de los filamentos RAD51/DNA, mientras CX3 actuaría en una fase posterior, estimulando el proceso de recombinación (Chun et al, 2013).

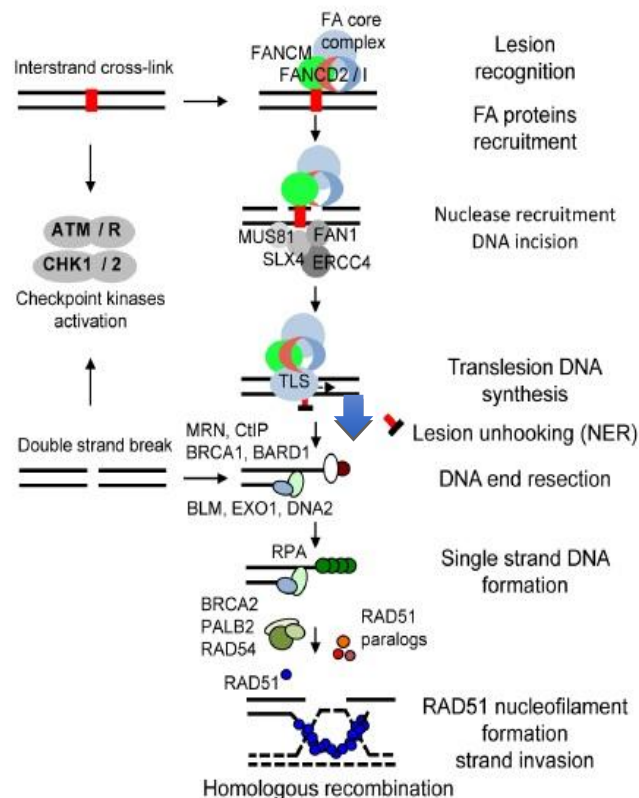
### **2.5. Muchos de los genes implicados en susceptibilidad genética al cáncer de mama y ovario, incluyendo *BRCA1*, *BRCA2*, *PALB2*, *RAD51C* y *RAD51D*, participan en la denominada ruta FA/BRCA, esencial para la reparación por recombinación homóloga de distintas lesiones en el DNA**

Probablemente, las lesiones del DNA potencialmente más dañinas para la célula son los enlaces covalentes entre hebras monocatenarias, que denominaremos ICLs (del inglés interstrand crosslinks). Estas lesiones, que físicamente impiden la replicación del DNA, se originan tanto por efectos no deseados de ciertos subproductos del metabolismo celular, como por agresiones externas, en particular muchos compuestos de Platino utilizados en quimioterapia (Ceccaldi et al., 2016). Otras lesiones del DNA, particularmente severas, son las roturas en la doble hebra del DNA. Estas lesiones aparecen cuando las dos cadenas de la hélice de DNA se rompen simultáneamente por dos lugares físicamente cercanos. Una reparación deficiente de estas lesiones puede provocar la delección o inserción de material genético, fragmentación cromosómica, translocaciones y pérdidas cromosómicas (Morgan et al., 1998).

Los procesos moleculares implicados en la reparación de ambos tipos de lesiones tienen diferencias significativas, en particular en los mecanismos de detección y procesamiento inicial de la lesión, pero tienen mecanismos de resolución comunes mediados por recombinación homóloga. De este modo, los mecanismos moleculares de reconocimiento y reparación de ICLs y roturas en la doble hebra del DNA convergen en lo que se denomina comúnmente ruta FA/BRCA.

Tal y como se muestra en la Figura 8, el denominado “core complex” del sistema FA, formado por 14 proteínas (FANCA, FANCB, FANCC, FANCE, FANCF, FANCG, FANCL, FANCM, FANCT, FAAP100, MHF1, MHF2, FAAP20 and FAAP24) participa en el reconocimiento inicial de los ICLs y en las primeras fases de su reparación. Por otro lado, los principales genes de susceptibilidad al cáncer de mama y/u ovario conocidos en la actualidad (genes donde se localizan variantes raras o muy raras de riesgo intermedio/elevado, como *BRCA1*, *BRCA2*, *CHEK2*, *PALB2*, *RAD51C*, *RAD51D*, *ATM*, y *BRIP1*) son fundamentales en el proceso de recombinación homóloga, común a la reparación de ICLs y roturas en la doble hebra del DNA que sólo puede operar en las fases S y G2 del ciclo celular (necesita de una segunda copia de DNA idéntica que actúe como molde para restaurar la información genética perdida de la cadena dañada). Aunque no se muestra en la Figura 8, la conexión entre los mecanismos moleculares que detectan/reparan ICLs y las roturas en la doble hebra del DNA son más profundos. Así, se ha descrito que: 1) *BRCA1* (también denominado *FANCS*), *BRCA2* (*FANCD1*) y *RAD51* (*FANCR*) participan en la estabilización de la horquilla de replicación bloqueada por una

ICL. 2) *FANCD2* colabora con *BRCA1* en las fases iniciales de la reparación de roturas en la doble hebra del DNA por recombinación homóloga; y 3) *BRIP1* (*FANCI*) participa en la recombinación homóloga, pero también en el proceso de síntesis de DNA denominado “*translation DNA synthesis*”, esencial para la reparación de ICLs (revisado en Konstantinopoulos et al., 2015, Ceccaldi et al, 2016).



**Figura 8.** Esquema de la ruta FA/BRCA, esencial para la reparación de lesiones ICLs y roturas en la doble hebra del DNA por recombinación homóloga (konstantinopoulos et al., 2015).

A nivel clínico, la estrecha relación molecular entre la reparación de ICLs y la recombinación homóloga se refleja en el hecho de que una serie de genes se asocian con dos manifestaciones clínicas bien distintas en función de que se hereden mutaciones de pérdida de función en los dos alelos o en un único alelo. En el primer caso, estamos ante un síndrome recesivo denominado Anemia de Fanconi caracterizado por supresión de la función de la médula ósea; presentar diversas malformaciones, predisposición a neoplasias hematológicas y tumores sólidos (D’Andrea and Grompe 2003; Joenje, H. & Patel, K. J, 2001).

En el segundo caso, ante un síndrome autosómico dominante de susceptibilidad al cáncer de mama y/u ovario. Este comportamiento clínico se ha observado asociado a *BRCA2*, *PALB2*, *BRIP1*, *RAD51C* y *BRCA1*, aunque las manifestaciones clínicas pueden ser sensiblemente distintas en función del gen concreto alterado. En el caso de *BRCA1* y *BRCA2*, se piensa que la

combinación de dos alelos nulos no es compatible con el desarrollo embrionario, y sólo determinados heterocigotos compuestos en los que (un alelo nulo combinado con un alelo hipomorfo) son capaces de asociarse con síndromes de Anemia de Fanconi más o menos atípicos (Ludwig, T et al., 1997).

### 3. Susceptibilidad Genética al cáncer de mama y ovario en la clínica: de *BRCA1* y *BRCA2* a los paneles genéticos

#### 3.1. Test genético en cáncer de mama y ovario familiar

De forma individual, los defectos genéticos en *BRCA1* y *BRCA2* son la contribución más significativa a la heredabilidad del cáncer de mama y ovario (Figura 2), y durante muchos años han sido los únicos factores genéticos utilizados en el asesoramiento de las familias con antecedentes de estas enfermedades. Sin embargo, hoy en día está bien establecido que los defectos genéticos en *ATM*, *PALB2* y *CHEK2* (cáncer de mama); y *RAD51C*, *RAD51D* y *BRIP1* (cáncer de ovario) tienen una contribución pequeña pero significativa a la heredabilidad de estas enfermedades (Figura 2), lo que justifica su inclusión en paneles multigénicos con finalidad diagnóstica (Easton et al, 2015). La mayoría de paneles con finalidad diagnóstica incluyen muchos otros genes (la mayoría relacionados directa o indirectamente con la ruta FA/BRCA, como por ejemplo *CHEK2*, *RAD50*, *RAD51B*, *XRCC2*, *BARD1*, *FANCM*, *FAM175A/Abraxas*, *SLX4*, *NBN/NBS1*, *MRE11A*, y otros) para los que ni siquiera está bien establecida una asociación con el riesgo de desarrollar cáncer de mama y ovario (Walsh et al, 2011, Kiiski et al, 2014; Kurian et al, 2014; Tung et al, 2015, revisado en Couch et al, 2014; Easton et al, 2015).

Probablemente, el desarrollo e implementación clínica de un modelo de susceptibilidad poligénico (combinado con factores no genéticos) como el esbozado en la sección 1.4, permitirá en un futuro cercano asignar un riesgo genético de desarrollar cáncer de mama y ovario a cualquier individuo de la población que desee testarse. Riesgo que será personal y esencialmente independiente de la historia familiar. Sin embargo, la realidad actual de los test genéticos con fines diagnóstico en cáncer de mama y ovario es mucho más prosaica, y casi sin excepciones, parte del supuesto de que el exceso de riesgo familiar es esencialmente monogénico. En este esquema simplificador, el test genético sólo tiene dos resultados posibles: **informativo** (se ha detectado una variante genética inequívocamente deletérea que asumimos que explica el síndrome familiar), y **no informativo** (no hemos

encontrado ninguna variante que de forma evidente nos permita explicar el síndrome familiar).

En el caso de un test informativo, la mutación identificada se utilizará para individualizar el asesoramiento genético en la familia. A los portadores se les atribuirá un riesgo de desarrollar cáncer de mama y/u ovario de acuerdo con las estimaciones existentes (Tabla 1), y a los no portadores se les atribuirá un riesgo equivalente al de la población general (Korde et al., 2011). En el caso de un test no informativo, no podremos individualizar el asesoramiento dentro de la familia.

En el caso de los test genéticos *BRCA1* y *BRCA2*, este esquema de trabajo ha demostrado beneficio clínico, reduciendo la incidencia y mortalidad por cáncer de mama y ovario (McCarthy A and Armstrong K, 2014). Hasta qué punto este esquema de trabajo puede ofrecer un beneficio clínico real en familias portadoras de variantes inequívocamente deletéreas en genes asociados a niveles de riesgo sensiblemente menores (Tabla 1), no está claro.

El ejemplo paradigmático de variante informativa en los genes supresores de tumores *BRCA1* y *BRCA2* (también en los genes *PALB2*, *RAD51C* y *RAD51D*) es una mutación sin sentido (*nonsense*), o una pequeña inserción/delección (*Indel*), que introduce un codón de parada prematuro (PTC, del inglés premature termination codon) y, en consecuencia, podemos inferir que inactiva el gen (y por tanto, asumir que se asocia con un determinado nivel de riesgo), sin necesidad de información adicional. La inmensa mayoría de mutaciones deletéreas en *BRCA1* y *BRCA2* son de este tipo, y probablemente lo mismo es cierto para *PALB2*, *RAD51C* y *RAD51D*, aunque en el caso de *RAD51C* se ha sugerido un exceso de mutaciones deletéreas del tipo *missense* (Osorio et al., 2012).

La identificación de los genes de susceptibilidad *BRCA1* y *BRCA2* transformó profundamente el manejo y el tipo de asesoramiento que podían recibir las familias con síndrome de cáncer de mama y ovario hereditario (HBOC). Por primera vez, se hizo posible distinguir los individuos portadores de los no portadores y ofrecer así un asesoramiento individualizado a cada miembro de la familia. Sin embargo, el consejo genético presenta una serie de implicaciones clínicas, éticas, psicológicas (y también económicas) que exigen una cuidadosa valoración de cada caso en particular. Tras una evaluación de la historia médica y familiar, se recomendará la realización de un test genético de *BRCA1/2* a aquellas familias con una probabilidad alta de que sufran un síndrome HBOC, y muy importante, aquellas familias donde la identificación de una variante genética inequívocamente causal pueda impactar positivamente en el manejo de la familia.



Además de seleccionar a las familias a las que se les ofrece la posibilidad de la realización de un test genético, también es importante la selección del individuo a testar, dentro de todos los miembros de una familia. El primer individuo de una familia donde se realiza un test se llama individuo **probando**. Su elección es clave para la interpretación del resultado del test genético, ya que únicamente se realiza un test genético por familia (en el individuo probando) y el resultado se aplica a todos sus familiares. Debe escogerse al individuo con mayor probabilidad de ser portador dentro de una familia; preferiblemente afecto (o portador obligatorio) y, a ser, posible, el de mayor afectación y menor edad al diagnóstico.

CM/CO	GEN	Método	RR (95% IC; <i>p</i> )	RA (edad)	Referencia
CM	<i>BRCA1</i>	Meta-análisis	11,4	75% (80 años)	Easton et al., 2015
	<i>BRCA2</i>	Meta-análisis	11,7	76% (80 años)	Easton et al., 2015
	<i>RAD51C</i>	Kin-Cohort/Segregación	0,9 (0,4-1,7)	-	Loveday et al., 2012
	<i>RAD51D</i>	Kin-Cohort/Segregación	1,3 (0,6-3,0)	-	Loveday et al., 2011
	<i>PALB2</i>	Meta-análisis	5,3 (3,0-9,4; $4 \times 10^{-10}$ )	45% (80 años)	Easton et al., 2015
CO	<i>BRCA1</i>	Estudio prospectivo	-	59% (70 años)	Mavaddat et al., 2013
	<i>BRCA2</i>	Estudio prospectivo	-	17% (70 años)	Mavaddat et al., 2013
	<i>RAD51C</i>	Kin-Cohort/Segregación	5,9 (2,9-11,9; $7.6 \times 10^{-7}$ )	>9% (80 años)	Loveday et al., 2012
	<i>RAD51D</i>	Kin-Cohort/segregación	6,3 (2,9-13,8; $4,8 \times 10^{-6}$ )	-	Loveday et al., 2011
	<i>PALB2</i>	Kin-Cohort/segregación	2,3 (0,8–7,0)	-	Antoniou et al., 2014

**Tabla 1. Algunas de las estimaciones de riesgo publicadas para *BRCA1*, *BRCA2*, *PALB2*, *RAD51C* y *RAD51D*.** La tabla resume estudios representativos con estimaciones de riesgo para cáncer de mama y cáncer de ovario. Se indica el método empleado para la determinación del riesgo y el resultado obtenido. Dependiendo del estudio, las estimaciones se ofrecen en forma de riesgo relativo (RR), o riesgo acumulado (RA) a los 70/80 años. Las mutaciones en *PALB2* confieren mayor riesgo de desarrollar cáncer de mama que las mutaciones en otros genes para los que existen estimaciones de riesgo razonablemente precisas (Easton et al, 2015), como *CHEK2* 1100delC (RR=3, RA-80 años =29% y *ATM* (RR=2,8; RA-80 años=27%). Las mutaciones en *PALB2* se asocian también con un riesgo muy elevado (RR=8,3; 0,8-88-6) de cáncer de mama en varón (Antoniou et al, 2014).

### 3.2. El problema de las variantes genéticas de significado clínico incierto

Una variante genética de significado clínico incierto (VUS, del inglés *variant of uncertain significance*) es un cambio en la secuencia genética para el que carecemos (en el momento de redactar nuestro informe de laboratorio) de datos que nos permitan concluir (o excluir) su relevancia clínica. Estos resultados además de no informativos añaden un alto grado de incertidumbre al resultado del test genético. Inevitablemente, el paso de los test *BRCA1* y *BRCA2* a los paneles multigénicos ha multiplicado el número de VUS identificadas en los test genéticos (Couch et al., 2014).

En general, las VUS son variantes genéticas raras (MAF  $<<0.01$ ) para las que, en el momento de emitir un informe genético, no nos es posible asumir o descartar un efecto deletéreo (ejemplos típicos son los cambios *missense*, los cambios sinónimos y la mayoría de cambios intrónicos).

Las VUS son un problema importante asociado a los test genéticos *BRCA*. En esencia, un test genético *BRCA* que concluye con la identificación de una VUS, es un test **no informativo** exactamente igual que cualquier otro test donde no se identifican mutaciones deletéreas. Sin embargo, la identificación de variantes sin clasificar nunca es deseable y añade complicaciones adicionales a un proceso de asesoramiento genético ya de por sí, complejo (Eccles et al., 2015).

Por un lado, es difícil transmitir de forma adecuada la información al portador sin incrementar innecesariamente el estrés emocional. Es frecuente que el portador concluya de forma errónea que la VUS es con toda probabilidad la causante del cáncer en la familia (la realidad es bien distinta, lo más probable es que se trata simplemente de una variante genética rara sin implicaciones clínicas).

Por otro lado, si queremos contribuir a “clasificar clínicamente la variante”, es probable que necesitemos involucrar a familiares para intentar hacer estudios de co-segregación, para recabar información clínica adicional, informes de anatomía patológica, solicitar nueva extracción de sangre para aislar RNA, etc. Todo ello, muchas veces no contribuye más que a extender el estrés emocional a otros familiares sin aportar (en la mayoría de los casos) ningún beneficio concreto a la familia.

Finalmente, a todo ello hay que sumar un problema importante de falta de estandarización en la clasificación clínica de variantes. En general, distintas variantes genéticas se asocian con distintos niveles de evidencia apoyando o descartando su potencial relevancia clínica. Además, las guías o documentos consenso disponibles en la actualidad están lejos de resolver todos los casos. El resultado final es que una misma variante genética pueda asociarse

a clasificaciones clínicas distintas (en algunos casos abiertamente contradictorias) en diferentes bases de datos y/o informes de laboratorio.

A pesar de la falta de estandarización en relación al *cómo* clasificar las variantes, existen aspectos relacionados con la clasificación clínica de variantes genéticas que hoy en día sí gozan de amplio consenso entre los profesionales, en particular en lo que se refiere a la comunicación y manejo de dichas variantes una vez clasificadas. Así, se recomienda clasificar clínicamente las variantes genéticas de acuerdo con un sistema de 5 categorías asociado a recomendaciones clínicas (Tabla 2), propuesto inicialmente por un panel de expertos de la IARC (*International Agency for Research on Cancer*). Según este esquema (Plon S et al., 2008), las variantes Clase-1 no tienen relevancia clínica (probabilidad de ser patogénicas,  $p < 0.001$ ), las variantes Clase-5 son patogénicas ( $p > 0.99$ ), y las variantes Clase-2, -3, y -4 se encuentran en situaciones intermedias.

Las variantes Clase-3 engloban un rango de probabilidades muy amplio (0.05-0.99) y, estrictamente hablando, equivalen a las VUSs (Eccles DM et al., 2015).

Clase	Definición Cuantitativa (P)	Definición Cualitativa	¿Incluir la variante en el informe de laboratorio?	¿Test Genético <u>predictivo</u> en familiares?	¿Test Genético con fines científicos en familiares?	¿Seguimiento?
Clase-5	>0.99	Patogénica	Sí	Sí	No	Test Informativo. Medidas de seguimiento y/o profilaxis correspondientes a individuos de alto riesgo
Clase-4	0.95-0.99	Probablemente Patogénica	Sí	Sí	Deseable	
Clase-3	0.05-0.949	Significado Clínico Desconocido	Sí	No	Deseable	Test No Informativo. Medidas basadas en la historia familiar y/o personal y en otros factores de riesgo.
Clase-2	0.001-0.049	Probablemente No Patogénica	¿Sí/No?	No	Deseable	
Clase-1	<0.001	No Patogénica	No	No	No	

**Tabla 2. Clasificación de las variantes genéticas propuesta por la agencia internacional para la investigación en cáncer (IARC)** (Tabla adaptada de Plon et al., 2008; Eccles et al., 2015).

Como se observa en la tabla 2, el sistema de clasificación en 5 categorías combina una descripción “cualitativa” con una descripción “cuantitativa”, expresada esta última en términos de probabilidad. Actualmente, la inmensa mayoría de las variantes genéticas en genes de susceptibilidad al cáncer de mama y ovario se realiza de forma cualitativa, y sólo una pequeña proporción de variantes genéticas en *BRCA1* y *BRCA2* se han podido clasificar de forma cuantitativa.

### 3.2.1. Clasificación Cualitativa

Este sistema de clasificación permite asignar a cada variante identificada en un test genético una clase clínica (Clase-1, -2, -3, -4 o -5) siguiendo algoritmos de clasificación que cuenten, a ser posible, con el consenso de un panel de expertos. En este sentido, el consorcio ENIGMA (Spurdle et al., 2012) ha desarrollado documentos que describen con mucho detalle un algoritmo para la clasificación de variantes genéticas en *BRCA1* y *BRCA2*.

Estos documentos se recogen en la dirección: <https://enigmaconsortium.org/library/general-documents/>. En general, en la presente tesis doctoral hemos clasificado las variantes de *BRCA1* y *BRCA2* de acuerdo al algoritmo del consorcio ENIGMA. La tabla 3 muestra algunos ejemplos de reglas cualitativas incorporadas en dicho algoritmo. Como se indica en el pie de figura, algunas reglas de clasificación no son más que el resultado de un consenso entre paneles de expertos; mientras otras, se apoyan en datos objetivos (la clasificación multifactorial de determinadas variantes individuales). Desafortunadamente, no se han desarrollado aún algoritmos equivalentes para la clasificación cualitativa de variantes genéticas en *RAD51C*, *RAD51D* y *PALB2*.

Clase	Definición Cualitativa	Regla de clasificación
5	Patogénica	➤ Variante que introduce un codón de parada prematuro en <i>BRCA2</i> antes del codón 3309 <sup>1</sup> .
4	Probablemente patogénica	➤ Variante que codifica el mismo cambio de aminoácido que una variante patogénica previamente establecida como Clase-5 <sup>2</sup>
3	Significado Clínico Incierto	➤ Variante que introduce un codón de parada prematuro en <i>BRCA2</i> entre los codones 3309 y 3325 <sup>1</sup>
2	Probablemente no patogénica	➤ Sustitución que codifica para el mismo cambio de aminoácido que una variante establecida previamente como Clase-1 (y no se predice bioinformáticamente que afecte al proceso de <i>splicing</i> ) <sup>2,3</sup>
1	No patogénica	➤ Cualquier variante con frecuencia >1% en 1000 genomas, EVS, o ExAc

**Tabla 3. Algunos ejemplos representativos de reglas cualitativas incluidas en el algoritmo ENIGMA de clasificación clínica de variantes genéticas en *BRCA1* y *BRCA2*.** En la tabla 4 y en la Discusión de nuestros resultados se mostrarán otros ejemplos de reglas cualitativas de clasificación. <sup>1</sup> Las reglas se basan en los resultados de la clasificación multifactorial de c.9924C>G (p.Tyr3308Ter), patogénica (datos pendientes de publicación), y c.9976A>T (p.Lys3326Ter), no patogénica (Meeks et al, 2015). Actualmente no hay información clínica disponible públicamente que apoye la patogenicidad de las variantes sin sentido (o de desplazamiento del marco de lectura) localizadas entre los codones 3309 y 3325. <sup>2</sup> Regla cualitativa basada exclusivamente en la opinión consensuada de un panel de expertos. <sup>3</sup> Es muy improbable que una variante *missense* de estas características tenga un efecto clínico similar al de una variante patogénica de alto riesgo (Vallee et al., 2016).

### 3.2.2. Clasificación Cuantitativa

Lógicamente, la descripción “cuantitativa” (mucho más precisa) sólo tiene sentido si existe un método para estimar la probabilidad de que una variante sea patogénica. Este método es el denominado *análisis multifactorial*.

Independientemente de su denominación, el análisis multifactorial parte de unos presupuestos esencialmente pragmáticos que se puede sintetizar del siguiente modo: dado que las variantes genéticas sin clasificar son por definición raras, ningún laboratorio individual tiene información suficiente para clasificarlas. Sin embargo, distintos laboratorios pueden tener evidencias parciales. Siempre que esas evidencias se puedan expresar en forma de cocientes de probabilidades (razón de verosimilitud o más comúnmente en inglés *likelihood ratios*), las distintas evidencias de los distintos laboratorios se podrán combinar (multiplicar) hasta alcanzar un nivel de evidencia suficiente.

El análisis multifactorial es el mejor método disponible en la actualidad para clasificar clínicamente las variantes genéticas, entre otras cosas porque al contrario de los ensayos funcionales o los análisis *in vitro* de *splicing*, no requiere ninguna asunción previa sobre el posible mecanismo de acción de la variante. Así, para un análisis multifactorial es irrelevante si una variante que definimos como *missense* es patogénica por introducir alteraciones estructurales en la proteína, por inducir un *splicing* aberrante, por modificar la estructura de la cromatina, o por disminuir los niveles de expresión del gen. En cualquier caso, será identificada correctamente como patogénica. Los análisis multifactoriales se desarrollaron inicialmente para la clasificación de variantes genéticas en *BRCA1* y *BRCA2* (Lindor NM et al., 2012; Goldgar DE et al., 2004), pero más recientemente se han desarrollado también análisis multifactoriales en genes MMR (Thompson BA et al., 2013).

Hasta donde sabemos, no se han desarrollado modelos multifactoriales para clasificar cuantitativamente variantes en *RAD51C*, *RAD51D* y *PALB2*.

Las evidencias más comúnmente incorporadas en los análisis multifactoriales actuales son la historia familiar, la segregación, datos histopatológicos y moleculares de los tumores (estatus del RE y grado histológico en el caso de *BRCA1* y *BRCA2*, y datos de pérdida de heterocigosidad, LOH). Sin embargo, el modelo permite incorporar cualquier otro tipo de evidencia, siempre y cuando se pueda expresar en forma de razón de probabilidades (y se refiera a variables independientes de las incorporadas previamente en el modelo).

Para ilustrar el concepto de evidencia expresada en forma de razón de probabilidades, vamos a utilizar un ejemplo muy simplificado relacionado con la clasificación de variantes en *BRCA1*. Imaginemos que tenemos información del receptor de estrógeno en 3 tumores de mama diagnosticados en portadores de una variante genética sin clasificar en *BRCA1*, y que 2 de estos 3 tumores son RE negativos. Supongamos también que tenemos una buena estimación de la probabilidad de que un tumor de mama sea RE negativo en población general (por ejemplo, 10%) y en portadores de mutación patogénica en *BRCA1* (por ejemplo, 70%). En estas condiciones, es evidente que la identificación de un 66% de tumores RE negativos es más probable que ocurra si la variante en *BRCA1* es una verdadera mutación patogénica que si no lo es, en concreto es  $\approx 16$  veces más probable ( $p=0,147$  vs.  $p=0,009$ ). Si somos capaces de recolectar información de 3 tumores adicionales y la proporción de tumores RE negativo se mantiene en el 66% (4 de 6), la evidencia a favor de que la mutación sea patogénica aumentará de forma sustancial. En concreto, ahora es  $\approx 267$  veces más probable ( $p=0.021609$  vs.  $p=0.000081$ ). Como es lógico, la probabilidad no se ve sólo afectada por los datos referentes a la variante en concreto, sino también por los valores de referencia que introducimos en el modelo. Si, por ejemplo, en un nuevo estudio epidemiológico se llega a la conclusión de que las mejores estimaciones para la probabilidad de que un tumor de mama sea RE negativo es del 15% en población general (en lugar del 10%) y del 60% en portadores de mutación patogénica en *BRCA1* (en lugar del 70%), es fácil comprobar que los mismos datos que poseemos en relación a nuestra variante (4 de 6 tumores RE negativo) nos generarán una probabilidad de que la variante patogénica sea sustancialmente menor.

Dicho de otro modo, al aproximarse los valores de referencia para el estatus RE negativo en población general y en portadores de mutación patogénica, el poder discriminatorio disminuye.

### **3.3. Un análisis de *Splicing* puede aportar información importante para la clasificación clínica de variantes genéticas en *BRCA1* y *BRCA2***

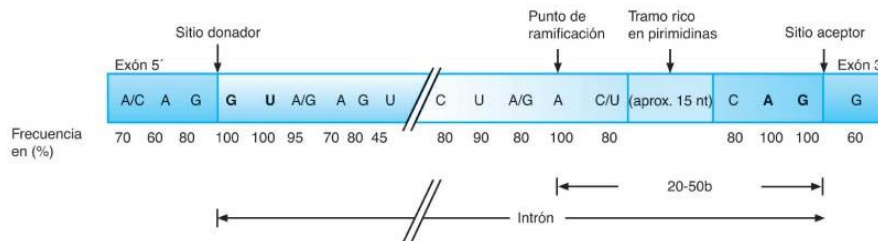
El procesamiento correcto de los intrones es un requisito imprescindible para la síntesis de una proteína. Variables genéticas que producen defectos en el *splicing* se han demostrado como causa de múltiples enfermedades hereditarias (Wang et al., 2012).

El *Splicing* es un proceso bioquímico complejo por el cual se eliminan las secuencias intrónicas de los RNA nacientes pre-RNA<sub>m</sub> y las secuencias exónicas flanqueantes se unen de



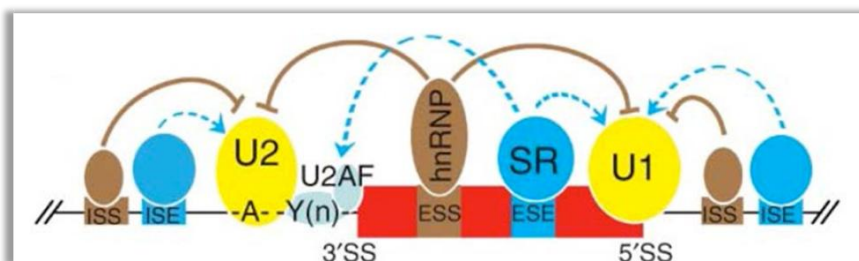
forma secuencial para dar lugar a un RNAm maduro (Gilbert, 1978). Para ello, un complejo riboproteico de conformación dinámica llamado espliceosoma interacciona con el pre-RNAm, reconoce las secuencias intrónicas y cataliza la doble reacción de transesterificación que caracteriza bioquímicamente al proceso de *splicing*. Se han descrito dos espliceosomas (mayor y menor) que tienen distinta composición de riboproteínas y reconocen distintos tipos de intrones (Hegele et al., 2012; Turunen et al., 2012). La mayoría de intrones son reconocidos y eliminados por el espliceosoma mayor.

El reconocimiento de las secuencias intrónicas que deben ser eliminadas del pre-RNAm se basa en la presencia de cuatro pequeñas secuencias intrónicas (donador y aceptor de *splicing*, punto de ramificación, y tracto rico en pirimidinas denominadas generalmente secuencias básicas de *splicing*). Los extremos de los exones, en particular el extremo 3' también son relevantes en el reconocimiento de los límites exón-intrón (Figura 9).



**Figura 9. Estructura de los límites intrón-exón (secuencias básicas de *splicing*).** Cuatro señales son relevantes para definir estos límites. Como se puede observar, en general el grado de conservación de estas secuencias no es muy elevado. Las únicas regiones conservadas completamente son las primeras/últimas 2 pb del intrón. Estas secuencias se conocen como sitios canónicos (adaptado de Manual Moderno, <http://www.manualmoderno.com>).

Existen muchas otras proteínas que, reconociendo en el pre-RNAm otro tipo de señales (secuencias reguladoras del *splicing*), interaccionan con el espliceosoma, regulando positiva o negativamente su actividad (Wang y Burge, 2008); y por tanto, regulando el nivel de inclusión de determinados exones en el RNAm maduro (Figura 10).



**Figura 10. Secuencias reguladoras que median el proceso de *splicing*.** En azul se muestran secuencias que estimulan el *splicing* (ESE e ISE) a través de su unión con proteínas del tipo SR y que facilitan el reconocimiento de los sitios de *splicing* por parte de las subunidades del espliceosoma. En marrón, secuencias (ISS y ESS) que impiden el reconocimiento de los sitios de *splicing* (Lu Z-X et al., 2011).

Las secuencias básicas y reguladoras de *splicing* presentan características muy distintivas que tienen implicaciones en la clasificación clínica de variantes genéticas:

- i) Las secuencias básicas ocupan unas regiones muy concretas del gen (los límites exón/intrón) y están relativamente poco degeneradas (las posiciones intrónicas IVS  $\pm 1,2$  son invariantes). Este hecho se traduce en que se hayan desarrollado numerosos programas *in silico* que predicen con mucha precisión el efecto (negativo o positivo) que una variante puede tener sobre estas secuencias. Además, podremos asegurar, sin necesidad de realizar un estudio a nivel de RNA, que una variante que afecte a las posiciones invariantes IVS  $\pm 1,2$  afectará al *splicing* (en este hecho se basa la clasificación clínica Clase-4 para las variantes IVS  $\pm 1,2$  que propone el algoritmo de clasificación cualitativa de ENIGMA. Tabla 3).
- ii) Las secuencias reguladoras, por el contrario, están muy degeneradas, por lo que las predicciones *in silico* del efecto de una variante genética sobre una secuencia reguladora son muy imprecisas y no tienen ninguna aplicabilidad clínica. Por otro lado, estas señales se localizan en posiciones bien definidas de exones y/o intrones. De este modo, cualquier variante genética intrónica o exónica podría afectar a una de estas señales (y por tanto al *splicing*). Esta es una de las razones principales por las que el algoritmo de clasificación cualitativa de ENIGMA considera que las variantes sinónimas e intrónicas deben considerarse Clase-3.

El análisis *in vitro* del efecto sobre el *splicing* puede aportar información relevante para la clasificación cualitativa de una UVS (Spurdle et al., 2008). En concreto, el algoritmo de clasificación elaborado por el consorcio ENIGMA (que seguimos en esta tesis doctoral) considera los siguientes supuestos:

- i) Cualquier variante genética, independientemente de su localización, puede considerarse Clase-5 si un estudio *in vitro* demuestra que el alelo portador de la variante sólo produce mensajeros aberrantes incapaces de codificar proteínas funcionales. Un ejemplo clásico es una variante que provoca *skipping* de un exón que codifica para un dominio funcionalmente relevante, o *skipping* de un exón que introduce un desplazamiento del marco de lectura.
- ii) Cualquier variante sinónima (o intrónica fuera de los sitios consenso de *splicing*) rara (MAF<0.01) debe considerarse Clase-3 en ausencia de otras evidencias. Sin embargo, si un estudio *in vitro* demuestra que no hay alteración de *splicing*, estas variantes pueden reclasificarse como Clase-1.

- i) Una variante en un sitio consenso de *splicing* se considerará, en ausencia de estudio *in vitro* de *splicing*, como Clase-4, y sólo un estudio de *splicing* que demuestre la producción de mensajeros alterados, que producen desplazamientos del marco de lectura (o la pérdida de un dominio funcionalmente relevante), permitirán considerar una variante Clase-5. Un estudio de *splicing* con otro resultado puede obligar a reclasificar la variante en sitio consenso de *splicing* como Clase-3.

Por supuesto, para que un estudio *in vitro* de *splicing* pueda determinar la clasificación de una variante genética, debe cumplir con unos requisitos mínimos de calidad e interpretarse de forma adecuada (Whiley et al, 2014; Walker et al, 2013). Así, un estudio positivo (se demuestra una alteración de *splicing* asociada a una variante genética concreta) debe ser capaz también de demostrar/excluir la producción de mensajeros normales por parte del alelo de la variante. Del mismo modo, un estudio negativo (no se observa alteración de *splicing*) debe demostrar que estamos detectando expresión bi-alélica. Muchos estudios *in vitro* de *splicing* con finalidad diagnóstica en *BRCA1* y *BRCA2*, y otros genes de susceptibilidad al cáncer se han realizado a partir de RNA de portadores (Walker et al., 2013). Sin embargo, muchos de estos estudios no son concluyentes, ya que no siempre es posible confirmar expresión bi-alélica en un estudio negativo, o en el caso de un estudio positivo excluir la producción de mensajeros normales por parte del alelo de la variante. Estas limitaciones (derivadas del hecho de que en RNA de portadores tenemos la contaminación derivada de la expresión del alelo silvestre) se pueden superar realizando estudios *in vitro* de *splicing* en minigenes. Sin embargo, los minigenes no reproducen fielmente el contexto genómico de los exones, ni los niveles de expresión nativos. Por ello, siempre que es posible, es preferible trabajar en paralelo con RNA de portadores para poder descartar artefactos generados por la propia naturaleza del minigen (Acedo et al., 2012; Fraile-Bethencourt E et al., 2017).

Como se discutirá en la siguiente sección y en la discusión, la posible existencia de *splicing* alternativo debe ser tenida en cuenta a la hora de diseñar e interpretar tanto los estudios *in vitro* de *splicing* en portadores como los estudios *ex vivo* basados en minigenes.

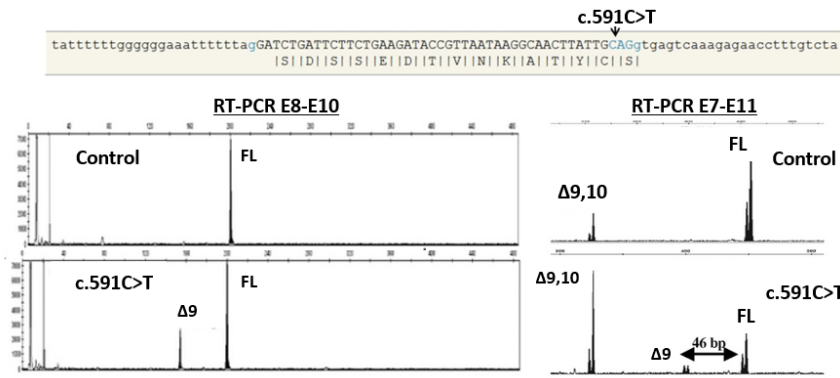
### 3.4. *Se ha demostrado que el *splicing* alternativo proporciona información potencialmente útil para la clasificación clínica de variantes genéticas en *BRCA1* y *BRCA2**

La inmensa mayoría de los genes humanos *clásicos* (codificadores de proteínas) son capaces de producir múltiples RNAm maduros (isoformas) vía mecanismos de *splicing* alternativo (Djebali et al, 2012). Se han descrito mecanismos de *splicing* alternativo en todos los metazoos, pero son más frecuentes en pájaros y mamíferos, organismos en los que además se observa una elevada tasa de *splicing* alternativo específico de tejido. Todo ello sugiere una relación estrecha entre *splicing* alternativo y complejidad fenotípica, aunque lo cierto es que las funciones específicas de la mayoría de isoformas se desconoce, y muchas de ellas carecen de regiones codificadoras de proteínas, o introducen codones de parada prematuros que presumiblemente inducen degradación selectiva (mecanismo NMD). Todo ello ha llevado a proponer que el *splicing* alternativo no sólo incrementa la complejidad del transcriptoma, sino que puede tener funciones reguladoras de la expresión, o representar ruido en el proceso de *splicing* (Mudge et al., 2011; Mudge et al., 2013).

De acuerdo con GENCODEv7, un gen humano *clásico* produce, de media, 6.31 isoformas por *splicing* alternativo, y los genes con más de 20 isoformas anotadas son raros (Djebali et al, 2012). Sin embargo, estos datos provienen de estudios pangenómicos que probablemente infravaloran el nivel real de *splicing* alternativo, como sugieren estudios de *splicing* alternativo gen-específicos que detectan decenas o cientos de isoformas en genes como *TP53*, *HOX*, *BRCA1* y *BRCA2* (Mercer et al., 2012; Colombo et al., 2014; Fackenthal et al., 2016).

Independientemente de su significado biológico, el *splicing* alternativo es un factor de enorme relevancia a la hora de definir las secuencias genómicas diana, anotar las variantes encontradas, e interpretar clínicamente los resultados que, sin embargo, tiende a ser ignorado en el diseño de la mayoría de los test genéticos. Así, sólo la descripción detallada del *splicing* (y su posible regulación tejido específico) permite definir de forma adecuada las secuencias diana de interés que deben incluirse en los test genéticos. En este sentido es importante recordar que, por ejemplo, determinadas regiones exónicas de *BRCA1* se identificaron mucho después de que los test genéticos en *BRCA1* se generalizaran (ElShamy et al, 2004; Fortin et al, 2005), y raramente son incluidas en dichos test. Por otro lado, sólo la descripción detallada del *splicing* permite anotar correctamente las variantes como exónicas (presentes en todos los RNAm maduros), intrónicas (excluidas de todos los RNAm) o mixtas (presentes en una fracción de los RNAm maduros). A su vez, estas anotaciones pueden ser tejido específicas. Es evidente que

dichas anotaciones pueden ser fundamentales para proceder a la interpretación funcional y clínica de las variantes genéticas. A su vez, el conocimiento detallado del *splicing* alternativo es fundamental para diseñar, analizar e interpretar de forma correcta los estudios *in vitro* de *splicing* (Dosil et al., 2010; Whiley et al., 2014). En concreto, en estos estudios una de las cuestiones de diseño más relevantes es la localización de los oligonucleótidos con los que se pretende realizar el experimento de RT-PCR (o los exones que deben incluirse en un minigen). El diseño experimental debe capturar todas las isoformas de *splicing* alternativo relevantes en función de la variante que se quiera interrogar. La importancia de estas consideraciones se ilustra perfectamente con el ejemplo de la variante *BRCA1* c.591C>T (p.=) que se ilustra en la Figura 11. Los análisis *in silico* predicen para esta variante, localizada en el extremo 3' del exón 9 de *BRCA1*, alta probabilidad de afectar al *splicing*. Un estudio *in vitro* mediante RT-PCR con oligonucleótidos localizados en los exones 8 y 10, confirma que la variante efectivamente afecta al *splicing*, causando la pérdida del exón 9 ( $\Delta 9$ ), una alteración que no conserva el marco de lectura; y por tanto, inactiva la función normal del gen. Este resultado podría llevar a considerar esta variante clínicamente relevante (patogénica) en el asesoramiento de una familia portadora. Un estudio equivalente, localizando los oligonucleótidos en los exones 7 y 11 muestra resultados de interpretación clínica muy distinta. La variante *BRCA1* c.591C>T (p.=) provoca *skipping* del exón 9, pero el efecto mayoritario es la pérdida de los exones 9 y 10 (o, dicho de otro modo, un aumento de la contribución relativa de la isoforma  $\Delta 9,10$ , presente en controles, a la expresión global. A diferencia de  $\Delta 9$ ;  $\Delta 9,10$  mantiene la fase de lectura. En este caso, el resultado del estudio *in vitro* no tiene una interpretación clínica obvia (¿el mensajero  $\Delta 9,10$  codifica una proteína funcional?) y nos vemos obligados a proponer una clasificación clínica 3 (variante de significado clínico desconocido). Un estudio multifactorial reciente (de la Hoya et al, 2016) ha demostrado que la variante no se asocia con ningún incremento en el riesgo de desarrollar cáncer de mama y ovario; y por tanto, no tienen ninguna relevancia clínica (Clase-1). En este caso concreto, el conocimiento previo de la existencia de la isoforma  $\Delta 9,10$  es crítico para seleccionar una combinación de oligonucleótidos adecuada.



**Figura 11. Análisis *in vitro* realizado por RT-PCR/EC.** La imagen superior izquierda muestra los resultados obtenidos en una RT-PCR comprendida entre los exones 8 y 10 de *BRCA1*. Mientras que las imágenes de la derecha muestran los resultados de una RT-PCR entre los exones 7 y 11 en muestras control y portadores de la variante de *BRCA1* c.591C>T. Figura adaptada de Dosil et al., 2012.

Las reglas de clasificación cualitativa del consorcio ENIGMA ya comentadas (ver 3.2.2) reconocen explícitamente la relevancia del *splicing* alternativo en el proceso de clasificación de variantes genéticas. En concreto, dichas reglas consideran que una variante genética localizada en un sitio consenso de *splicing* (variantes IVS±1,2), para la que no existe un estudio *in vitro* de *splicing*, debe considerarse probablemente patogénica (Clase-4); pero incluyen excepciones, determinados donadores/aceptores de *splicing* donde debemos ser más conservadores y considerar dichas variantes de significado clínico incierto (Clase-3) en base a la existencia de *splicing* alternativo potencialmente funcional, tal y como se detalla en la tabla 4.

Clase	Definición Cualitativa	Regla de clasificación
4	Probablemente patogénica	➤ Cualquier variante IVS±1,2 sin estudio <i>in vitro</i> de <i>splicing</i> , excepto las que deben considerarse Clase-3, tal y como se detalla a continuación
3	Significado Clínico Incierto	<p>Cualquier variante sin estudio <i>in vitro</i> de <i>splicing</i> que se localice en las siguientes posiciones:</p> <p><i>BRCA1</i> c.442-1,2 (IVS7-1,2)<sup>1</sup>  <i>BRCA1</i> c.548-1,2 (IVS8-1,2)<sup>2</sup>  <i>BRCA1</i> c.593+1,+2 (IVS9+1,+2)<sup>2</sup>  <i>BRCA1</i> c.594-1,-2 (IVS9-1,-2)<sup>2</sup>  <i>BRCA1</i> c.594-2 (IVS9-2)<sup>2</sup>  <i>BRCA1</i> c.670+1 (IVS10+1)<sup>2</sup>  <i>BRCA1</i> c.670+2 (IVS10+2)<sup>2</sup>  <i>BRCA1</i> c.4186-1 (IVS12-1)<sup>1</sup>  <i>BRCA1</i> c.4186-2 (IVS12-2)<sup>1</sup>  <i>BRCA1</i> c.4358-1 (IVS13-1)<sup>1</sup>  <i>BRCA1</i> c.4358-2 (IVS13-2)<sup>1</sup>  <i>BRCA2</i> c.6842-1 (IVS11-1)<sup>3</sup>  <i>BRCA2</i> c.6842-2 (IVS11-2)<sup>3</sup>  <i>BRCA2</i> c.6937+1 (IVS12+1)<sup>3</sup>  <i>BRCA2</i> c.6937+2 (IVS12+2)<sup>3</sup></p>

**Tabla 4. Algunos ejemplos de reglas cualitativas incluidas en el algoritmo ENIGMA de clasificación clínica de variantes genéticas en *BRCA1* y *BRCA2*<sup>1</sup>.** Los exones 8, 13 y 14 de *BRCA1* contienen dos aceptores de *splicing* en tandem (NAGNAG) funcionales, tal y como demuestra la identificación en controles de las isoformas *BRCA1* Δ8p<sub>3</sub>, Δ13p<sub>3</sub>, y Δ14p<sub>3</sub> (Colombo et al., 2014). Es previsible que las variantes incluidas en la tabla inactiven el aceptor proximal, pero no el distal, causando únicamente una pérdida de 3 nucleótidos en la secuencia codificante, lo que justifica la recomendación de considerarlas Clase-3. Hasta donde sabemos, no hay estudios experimentales de *splicing* para ninguna de estas variantes. <sup>2</sup> Es previsible que los portadores de variantes en estas posiciones produzcan niveles normales (o aumentados) de *BRCA1* Δ (9,10), un evento de *splicing* alternativo muy abundante que mantiene la fase de lectura y no elimina ningún dominio funcionalmente relevante conocido (Colombo et al., 2014), lo que justifica la recomendación Clase-3. Los resultados experimentales disponibles sustentan esta predicción (Dosil et al, 2010, de la Hoya et al, 2016, resultados de la presente tesis doctoral). <sup>3</sup> Es previsible que estas variantes causen *skipping* del exón 12. *BRCA2* Δ12 es una isoforma de *BRCA2* que se expresa de forma natural y mantiene la fase de lectura (Fackenthal et al., 2016). El exón 12 de *BRCA2* es funcionalmente redundante (Li et al., 2009) lo que justifica la recomendación Clase-3. Es importante recalcar que esta tabla no descarta de modo alguno que las variantes mencionadas sean patogénicas. Simplemente señala variantes donde es necesario realizar estudios *in vitro* de *splicing* y/o estimaciones directas de riesgo antes de asumir que la variante es patogénica.

# Objetivos





# OBJETIVOS

---

1. Demostrar la utilidad e impacto de los estudios de *splicing* en la clasificación de variantes en los genes *BRCA1* y *BRCA2*.
2. Caracterizar el *splicing* alternativo en los genes de susceptibilidad al cáncer de mama y ovario: *RAD51C*, *RAD51D* y *PALB2*, identificando eventos de *splicing* alternativo de posible relevancia clínica.



# Materiales y Métodos



# MATERIALES Y MÉTODOS

---

## 1. Obtención de RNAs y síntesis de DNA complementario

### 1.1. RNA de líneas Celulares linfoblastoides (LCLs)

Todas las LCLs utilizadas en la presente tesis doctoral se han generado y cultivado en centros de investigación con los que el laboratorio de oncología molecular del Hospital Clínico San Carlos (HCSC) ha establecido diversas colaboraciones científicas a través del consorcio internacional de investigación ENIGMA. Tal y como se indicará en resultados, dependiendo del experimento se han empleado LCLs sin tratar, o tratadas con Puromicina (un inhibidor del mecanismo NMD, del inglés *Non-sense Mediated Decay*). En algunos casos, el laboratorio colaborador nos ha enviado el RNA ya purificado. En otros casos, el laboratorio nos ha mandado las células cultivadas, y la extracción y purificación de RNA se ha realizado en nuestro laboratorio. En estos casos, la extracción de RNA se ha realizado con el sistema automático Magna Pure (Roche) de acuerdo a las instrucciones del fabricante.

### 1.2. RNA obtenido de células nucleadas de sangre periférica (Sangre)

En todos los casos, la extracción de RNA se realizó a partir de sangre periférica preservada en tubos Tempus (ThermoFisher). La purificación del RNA se realizó con el kit de extracción manual de ThermoFisher, siguiendo el protocolo especificado por el fabricante.

### 1.3. RNA de tejidos de Mama y Ovario Humanos Normales (No Cancerosos)

Se han empleado RNAs comerciales obtenidos de un tejido mamario normal (perteneciente a una mujer de origen Caucasiano fallecida de muerte súbita a los 27 años, Clontech 636576) y un tejido de ovario normal (*pool* generado a partir de tejido ovárico normal obtenido de 3 mujeres de origen caucasiano fallecidas de causa desconocida a los 40, 55 y 71 años, Clontech 636555). En ambos casos el RNA total se purificó mediante un protocolo basado en el uso de ticionato de guanidinio.

## 2. Cuantificación de RNA

La cuantificación de ácido ribonucleico (RNA) se realizó mediante espectrofotometría utilizando un equipo Nanodrop 1000 system (Thermo Scientific) de acuerdo a las recomendaciones del fabricante.

## 3. Síntesis de DNA complementario

La retrotranscripción se ha realizado con el Kit Prime Script RT Reagent Ki (Perfect Real Time) de TAKARA Biotechnologies, siguiendo las instrucciones del fabricante. En todos los casos se utilizaron como cebadores una mezcla de Oligo dT y Random-6-mers.

La reacción se realizó en un termociclador GeneAmp PCR system 2700 (Lifetechnologies) siguiendo el programa: extensión de 15' a 37°C, seguida de inactivación de la retrotranscriptasa a 85°C durante 5''.

Reactivos	Volumen
Tampón 5X Prime Script	2µl
RNA molde (<1µg)	*
PrimeScript RT Enzyme Mix I	0.5µl
Oligo dT Primer	0.5µl
Random 6 mers	0.5µl
dH2O (libre de RNAsas)	*
Volumen final	10µl

Tabla 5. Mezcla de la reacción RT-PCR

\*El volumen final dependerá de la concentración del RNA de partida. (El kit permite retro-transcribir una cantidad máxima de 1µg de RNA por reacción). Según las necesidades podemos escalar la reacción a un volumen final de 20 µl con resultados equivalentes.

### A. Estudios *in vitro* de *splicing*.

#### 1. SELECCIÓN DE MENSAJEROS: Transcritos de referencia

Para caracterizar, anotar y designar eventos de *splicing* alternativo en un determinado gen (o alteraciones de *splicing* inducidas por determinadas variantes genéticas en dicho gen) es indispensable definir previamente un mensajero de referencia para dicho gen, que no necesariamente es el mensajero de mayor relevancia biológica.

Para definir los mensajeros de referencia de los genes *BRCA1*, *BRCA2*, *RAD51C*, *RAD51D*, y *PALB2* hemos tenido en cuenta tanto los datos previos de la literatura científica y bases de datos de referencia, como la información de Ensembl y las predicciones realizadas por el sitio web APPRIS (<http://appris.bioinfo.cnio.es>).

Gen	Ensembl/APPRIS	RefSeq
<i>BRCA1</i>	ENST00000357654.7	NM_007294.3
<i>BRCA2</i>	ENST00000380152.7	NM_000059.3
<i>RAD51C</i>	ENST00000337432.8	NM_058216.2
<i>RAD51D</i>	ENST00000345365.10	NM_002878.3
	ENST00000590016.5	NM_001142571.1
<i>PALB2</i>	ENST00000261584.8	NM_024675.3

**Tabla 6. Mensajeros de referencia para cada gen estudiado.**

En el caso de *RAD51D*, la elección de un transcrito de referencia no es obvia, tal y como se analizará en resultados (ver sección 2, apartado 4.1). En el anexo I se detalla la selección del transcrito de referencia en relación a la información disponible en Ensembl.

## 2. Diseño cebadores

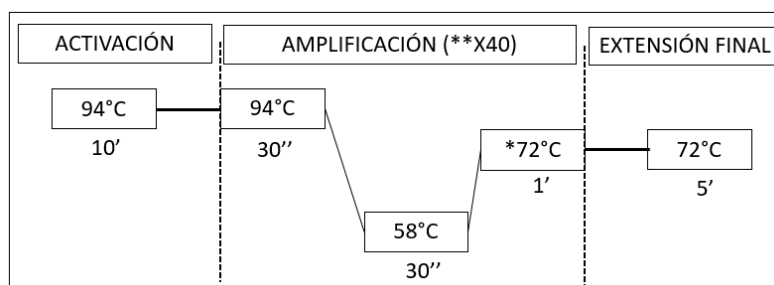
Todos los cebadores empleados para la amplificación de cDNA mediante PCRs han sido diseñados con la herramienta informática Primer3 input (v3.0.0) (<http://primer3plus.com/web3.0.0/primer3webinput.htm>).

La mayor parte de los cebadores antisense han sido marcados con un fluoróforo (FAM), con el fin de poder utilizarlos en electroforesis capilar. En algunos casos se han utilizado también cebadores *antisense* sin marcar para no interferir con reacciones de secuenciación Sanger. En el Anexo I se detalla la secuencia de todos los cebadores utilizados en los estudios de *splicing* para cada gen.

## 3. PCR

Todas las PCRs se han realizado a partir de un 1µl de cDNA, usando la enzima Fast Start Taq DNA polimerasa (ROCHE). Las reacciones se llevan a cabo en un termociclador GeneAmp PCR system 2700 (Lifetechnologies), siguiendo el programa:





\*En algunos casos, ha sido necesario emplear tiempos de elongación más largos para la amplificación preferencial de transcritos de tamaño mayor (>800pb). Se han empleado para esos casos tiempos de elongación de entre 120''-2'.

\*\*Para minimizar la amplificación preferencial de fragmentos pequeños e intentar mejorar la información cuantitativa, en algunos casos hemos reducido el número de ciclos de PCR a 33.

#### 4. Electroforesis en gel de Agarosa

Todas las amplificaciones se han comprobado con una electroforesis en gel de agarosa. Los geles utilizados tenían una concentración de agarosa del 2% (p/v) en la mayor parte de los casos. En los casos donde los productos obtenidos son mayores de 600 pares de bases (pb), se han empleado geles con una concentración de agarosa del 1% (p/v).

Para la preparación de los geles, la agarosa ha sido disuelta en tampón TBE 0,5X (Tris-Borato 45 mM pH 8,0; ácido bórico 45 mM, EDTA 1mM). Como agente intercalante se han añadido 2µl (por cada 100 ml) del agente Realsafe Nucleic Acid Staining solution (20.000x) de Real.

#### 5. Análisis de Fragmentos por Electroforesis Capilar

Tras su comprobación en agarosa, todos los productos amplificados en las RT-PCR han sido analizados por electroforesis capilar. Esta técnica nos permite analizar con muy elevada resolución ( $\pm 1,2$ nt), y alta sensibilidad los productos generados en cada una de las RT-PCRs.

Ese producto amplificado se cargó en una placa de 96 pocillos con una mezcla formada por 20µl de Formamida (Lifetechnologies) y 2µl de marcador Liz 500 (Lifetechnologies). En aquellos casos donde se esperaba un tamaño de fragmento de alrededor de 600pb, se ha empleado el LIZ600 (Lifetechnologies). El análisis se realizó en todos los casos en un secuenciador ABI Prism

3130 (Lifetechnologies) que montaba cuatro capilares de 50cm rellenos con polímero POP-7. Se emplearon condiciones de electroforesis estándar para este tipo de equipos (15Kw, 60°C), variando el tiempo/voltaje de inyección y el tiempo de carrera en función de la intensidad de señal obtenida en las PCRs y el tamaño de los fragmentos obtenidos.

Los resultados se analizan con el programa GeneMapper v4.0, que representa los datos de forma visual en electroferogramas, donde las abscisas representan el tamaño de los fragmentos en nucleótidos y las ordenadas la intensidad de fluorescencia en unidades arbitrarias.

Al analizar por electroforesis capilar productos de RT-PCR, esperamos identificar al menos un pico de tamaño compatible con el mensajero de referencia. Los posibles picos adicionales se consideran mensajeros adicionales: ya sean isoformas que se producen de forma natural por *splicing* alternativo en todas las muestras, o mensajeros aberrantes inducidos por la presencia de alguna variante genética.

Comparando el tamaño del pico correspondiente al mensajero de referencia con el tamaño de los picos adicionales, y teniendo en cuenta la presencia de potenciales donadores GT y aceptores AG en la secuencia genómica de referencia, podemos hacer predicciones sobre la identidad de los mensajeros adicionales (anotaciones basadas en datos de electroforesis capilar). En los casos en los que no encontramos una explicación plausible para explicar el tamaño de los picos, lo indicamos con un signo de interrogación (?).

### 6. Secuenciación Sanger

En la reacción de secuencia se ha empleado la química del agente BigDye Terminator vs3.1 de Lifetechnologies de acuerdo con el protocolo del fabricante. Posteriormente, los productos de secuenciación han sido purificados con el Kit de purificación BigDye Terminator de Lifetechnologies para eliminar los terminadores sobrantes, y evitar posibles ruidos de fondo. Las muestras secuenciadas se han analizado en un secuenciador ABI 3130 (Lifetechnologies).

### B. Caracterización del *splicing* alternativo

Para la caracterización del *splicing* alternativo se partió de dos fuentes de datos primarias independientes: 1) Datos primarios de RNAseq proporcionados por un laboratorio colaborador en el marco del consorcio ENIGMA (en esencia, estos datos consistían en coordenadas genómicas de intrones obtenidas mediante STAR y TopHat), y 2) datos de RT-PCR analizados mediante electroforesis capilar y/o secuenciación Sanger.

#### 1. RNASeq

Para la realización de esta tesis doctoral hemos contado con datos de RNAseq proporcionados por laboratorio colaboradores dentro del consorcio ENIGMA.

Los datos de RNAseq en LCLs fueron proporcionados por el Dr. Logan Walker, de la Universidad de Otago (Nueva Zelanda). En síntesis, los datos proceden de la secuenciación del transcriptoma completo de una célula linfoblastoide derivada de una mujer control sana tratada y sin tratar con un inhibidor del mecanismo NMD (Cicloheximida).

Las librerías se construyeron a partir de RNA total utilizando el kit TruSeq Stranded total RNA sample preparation kit (Illumina, San Diego, CA) siguiendo las instrucciones del fabricante, y se secuenciaron en un HiSeq2000 (Illumina). Las lecturas se mapearon contra el genoma de referencia Homo\_sapiens.GRCh37.7 utilizando el alineador STAR con parámetros por defecto, excepto el tamaño máximo permitido para un intrón que se restringió a 100.000Kb. Las mismas lecturas también se mapearon en paralelo con el alineador TopHat. Nosotros partimos de los archivos STAR y TopHat que contienen las coordenadas genómicas de las uniones exón/exón (coordenadas genómicas de intrones) para realizar nuestras anotaciones manuales de eventos de *splicing* alternativo (ver resultados).

Los datos de RNAseq en tejido de mama y fimbria ovárica sanos fueron proporcionados por el Dr. Anders Kvist, de la Universidad de Lund (Suecia). Las muestras de RNA proceden de tejidos normales de mama obtenidos en la cirugía de 9 mujeres con cáncer de mama (no portadoras de mutación germinal en *BRCA1*, *BRCA2*, *TP53*, *PALB2*, *PTEN*, *CHEK2* y *CDH1*), y de tejidos de fimbria ovárica obtenidos de ooforectomías profilácticas realizadas en dos mujeres posmenopáusicas sanas no portadoras de mutación en *BRCA1* o *BRCA2*. Los métodos de extracción de RNA y generación de librerías con el kit TruSeq Stranded mRNA (Illumina, San Diego, CA) se han descrito previamente. Las librerías se secuenciaron en un

equipo Illumina HiSeq 2000 (paired end 2x75 nucleótidos). Las lecturas se mapearon contra el genoma de referencia Homo\_sapiens.GRCh37.7 utilizando los alineadores STAR y TopHat con parámetros por defecto.

### 2. Designación de eventos de *splicing* alternativo, anotación biológica (biotipo) y anotación funcional.

Para la designación de eventos de *splicing* alternativo nos hemos guiado por trabajos previos del consorcio ENIGMA (Colombo et al, 2014; Fackenthal et al, 2016). En síntesis, se combinarán símbolos de acuerdo a las siguientes convenciones:

- La pérdida de nucleótidos respecto al mensajero de referencia (pérdida parcial o completa de unos o varios exones) se denotará con  $\Delta$ .
- La ganancia de nucleótidos (inclusión en el mensajero maduro de secuencia intrón de referencia) se denotará con  $\nabla$ .
- Las pérdidas (o ganancias) que afecten al extremo 5' de un exón (aceptor de *splicing* alternativo) se especifican con la letra "p".
- Las pérdidas (o ganancias) que afecten al extremo 3' de un exón (donador de *splicing* alternativo) se especificarán con la letra "q".
- El tamaño (en nucleótidos) de la alteración se indicará con un **subíndice** en el caso de una pérdida, o con un **superíndice** en el caso de que sea una inserción. A excepción de que sea un valor obvio (pérdida de exones de referencia completos).
- Se utilizarán comas "," para designar dos exones adyacentes y guiones bajos "\_" para designar varios exones adyacentes.
- La existencia de mensajeros que combinen más de un evento de *splicing* alternativo se indicará con el símbolo "+".

La anotación biológica (biotipo) y funcional de los eventos de *splicing* alternativo la hemos realizado de acuerdo a GENCODE (Mudge et al, 2011).

A modo de ejemplo, un evento designado *RAD51C*  $\nabla 7p^{22}$  indica la utilización de un aceptor de *splicing* alternativo situado en el intrón 7 de *RAD51C*, a 22 nucleótidos del aceptor de referencia para el exón 7 de dicho gen. El evento se anota como *acceptor shift* (biotipo) y funcionalmente como PTC-NMD (una inserción de 22 nucleótidos desplaza el marco de lectura,

genera un codón de parada prematuro y se predice que induzca el mecanismo NMD de degradación selectiva). Un evento designado *RAD51C* Δ5 indica la no inclusión (*skipping*) del exón 5 de *RAD51C* en el mensajero maduro. El evento se anota como *cassette* (biotipo) y funcionalmente como *No FS* (la pérdida del exón 5 representa una pérdida de 132 nucleótidos en el mensajero maduro, que mantiene la fase de lectura).

## 3. Análisis Cuantitativo del *splicing* alternativo mediante PCR digital

### 3.1. Diseño ensayos TaqMan

Para la cuantificación de las diferentes isoformas de *splicing*, hemos empleado dos ensayos TaqMan marcados con fluoróforos distintos. Un ensayo FAM que reconoce la unión exón-exón interrogada (característica de la isoforma de interés), y un ensayo VIC que reconoce una unión exón-exón característica del mensajero de referencia. Hemos empleado tanto ensayos TaqMan prediseñados por la casa comercial Life Technology como ensayos TaqMan diseñados específicamente para nuestro estudio. En este último caso, el diseño se ha realizado utilizando la herramienta on-line de LifeTechnology ([www.lifetechnologies.com](http://www.lifetechnologies.com)). En la tabla 7 y en el anexo I se describen los ensayos TaqMan empleados.

GEN	Marcaje	exón/exón	Secuencia de la unión exón/exón	Referencia
<i>BRCA1</i>	FAM	8-11p	5'-CATTGAATTGG-CTGCTTGTGAATTT-3'	Hs01556198
<i>BRCA1</i>	FAM	9-10	5'-CCCACA-CTGCAATAAG-3'	B1-E9E10
<i>BRCA1</i>	VIC	23-24	5'-CAATGGGCTTCCATG-CAATTGGGCAG-3'	Hs01556193
<i>BRCA2</i>	FAM	2-4	5'-CAAAGCAG-GAAGGAATG-3'	B2_E2E4
<i>BRCA2</i>	VIC	3-4	5'-AATTAGACTTAG-GAAGGAATGTTCC-3'	Hs00609076
<i>BRCA2</i>	VIC	26-27	5'-GAAACAAGCTTCTG-ATGTCTTCTCC-3'	Hs00609073
<i>RAD51D</i>	FAM	E2-E6	5'-TCTTACAAG-GCAGAAGCTCTCCGGA-3'	Hs00979541_m1
<i>RAD51D</i>	FAM	E3 <sup>119</sup> -E4	5'-TGGCATTGGCAG-TCTTGATAAACTG-3'	Hs00979557_m1
<i>RAD51D</i>	VIC	E3 <sup>119</sup> -E4	5'-TGGCATTGGCAG-TCTTGATAAACTG-3'	Hs00979557_m1
<i>RAD51D</i>	VIC	E9-E10	5'-CTTCCCGACAG-CCAACAGGTTTCCA-3'	Hs00979545_g1
<i>RAD51D</i>	FAM	E3 <sup>179</sup> -E4	5'-TTCACG-TCTTGATAAACTGCTTGAT-3'	Hs00979563_m1
<i>PALB2</i>	VIC	E4-E5	5'-GTGAAAG-GGAAGAAAAGTCGTCATC-3'	Hs00954119_m1
<i>PALB2</i>	FAM	E3-E6	5'-AACACTCAG-ACTGAAACAGCAGAGC-3'	PALB2_E3E6
<i>PALB2</i>	FAM	E11-E12	5'-TTCTGAAATG-GGGCTTCTCTTATT-3	Hs00226617_m1

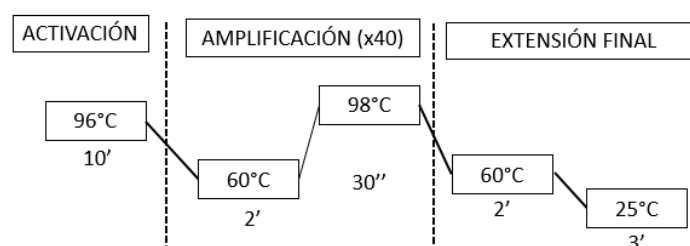
Tabla 7. Ensayos TaqMan empleados en la cuantificación por PCR digital.

### 3.2. Validación ensayos TaqMan

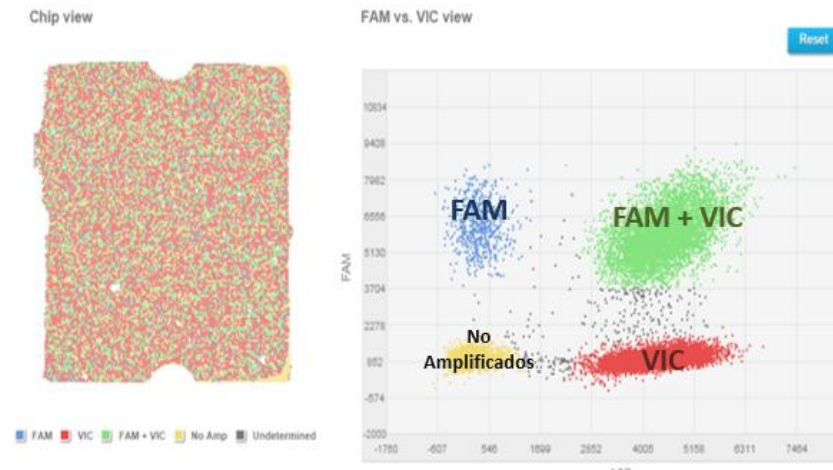
Los ensayos TaqMan diseñados para el análisis cuantitativo por PCR digital han sido validados mediante **PCR a tiempo Real**. Para evaluar el nivel de eficacia, los ensayos han sido validados por separado y en conjunto. Para la validación se han empleado cDNA de controles sanos de sangre que han sido cargados en una placa de 96 pocillos en combinación con 2µl de cada ensayo TaqMan. La amplificación se realizó usando el equipo 7500 Fast System (Life-technologies).

### 3.3. PCR digital

Las PCRs digitales se han realizado en un equipo QuantStudio 3D Digital PCR system (Life-technologies) utilizando chips 20k (20.000 particiones) y procediendo de acuerdo con el protocolo indicado por la casa comercial. En resumen, con ayuda de un brazo mecánico se siembra sobre cada chip 20k una mezcla formada por 7,2µl de una concentración adecuada de cDNA, 9µl de master-mix QuantStudio 3D, 0,9µl de sonda FAM y 0,9µl de sonda VIC. Si la concentración de partida de cDNA es correcta, de media cada una de las 20.000 particiones recibirá menos de una molécula de cDNA (es decir, en la mayoría de particiones habrá una o ninguna molécula de cDNA). A continuación, se cubre el chip con aceite de inmersión y se realiza una amplificación estándar por PCR.



Resulta importante llevar a cabo una carga eficaz de la muestra, pues de ello dependerá la separación de clusters (nube de puntos) que representan la expresión de las isoformas a cuantificar. Aunque son dos ensayos TaqMan los que se combinan en un mismo estudio, obtenemos cuatro poblaciones de expresión. La Figura 12 muestra un ejemplo representativo de este tipo de estudios.



**Figura 12. Calidad de la carga y separación de los clusters.** Se representa la eficacia con la que se ha repartido la muestra sobre las 20.000 particiones que comprenden el chip, y la eficiencia con la que se separan las distintas poblaciones de moléculas de cDNA. Como se puede observar, tenemos cuatro poblaciones distintas representadas con distintos colores, en función del tipo de molécula presente/ausente originalmente en cada una de las 20.000 particiones. En este ejemplo, la población de clusters **azul** corresponde a los mensajeros del evento de *splicing* alternativo  $\Delta 9,10$  (marcado con FAM). La población de clusters **rojo** representan al mensajero de referencia (marcado con VIC) reconoce la unión de los exones 23-24 que representa la expresión global de *BRCA1*. En **verde** se representa la población de clusters que reconocen mensajeros  $\Delta 9, 10$  y la unión E23-24. Y la población **amarilla** representa las moléculas ausentes o no amplificadas en las particiones.

# Resultados





# RESULTADOS

---

## I. Análisis de *splicing* y clasificación clínica de variantes genéticas en *BRCA1* y *BRCA2*

En nuestro centro, los análisis genéticos que se realizan sobre los genes *BRCA1/2* (Test *BRCA1/2*) son capaces de identificar variantes patogénicas en un escaso 18-20% de los casos estudiados (cifras similares a las de otros centros de referencia). Es decir, aproximadamente el 80% de los test que se realizan en estos genes no son informativos (no son capaces de identificar una variante genética inequívocamente causal, y por tanto no tienen ningún impacto en el manejo posterior de los individuos estudiados o sus familias).

Como se ha comentado en la introducción, los test no informativos incluyen aquellos estudios en los que sólo se identifican variantes consideradas Clase-1 (no patogénicas), pero también estudios en los que se identifican variantes Clase-2 (probablemente no patogénicas), y/o variantes Clase-3 (significado clínico incierto).

Desafortunadamente, el porcentaje de test BRCA que identifican variantes Clase-3 es relativamente elevado (5%-10%). Por ello, resulta urgente desarrollar estrategias que nos puedan ayudar a interpretar correctamente el significado clínico de este tipo de variantes. Como se verá a continuación, el análisis de *splicing* en RNA de portadores y controles mediante RT-PCR seguida de electroforesis capilar (RT-PCR/EC), secuenciación Sanger y/o PCR digital (dPCR) puede ser una de estas estrategias.

Dependiendo del tipo de variante estudiada y el resultado obtenido, este tipo de estudios nos permite:

- demostrar que una variante (sinónima e intrónica) no afecta al *splicing*, de tal modo que puede reclasificarse como Clase-1.
- identificar alteraciones de *splicing* asociadas a variantes Clase-1. En estos casos, el objeto del estudio es comprender mejor la relación entre determinadas alteraciones de *splicing* y el nivel de riesgo (tolerancia clínica a las aberraciones de *splicing*). En algunos casos, este tipo de análisis nos permitirá definir nuevas reglas cualitativas de clasificación de variantes.
- demostrar que una variante genética afecta al *splicing* de tal modo que puede reclasificarse como Clase-5.

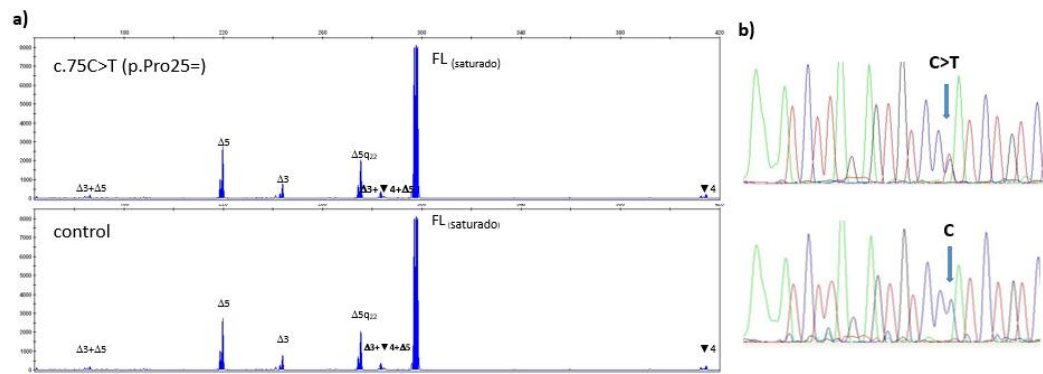
En la presente tesis doctoral se muestran los resultados del análisis de *splicing* realizado en portadores de 21 variantes genéticas en *BRCA1* y 17 variantes genéticas en *BRCA2*. Las tablas 8 y 9 resumen los resultados de los 38 estudios. A continuación, mostramos en detalle dos ejemplos representativos de estudios de *splicing* que nos han permitido reclasificar variantes (*BRCA1* c.5152+4A>T; *BRCA1* c.75C>T), y un ejemplo representativo de estudio de *splicing* realizado en una variante (*BRCA1* c.594-2A>C) con clasificación multifactorial previa (Clase-1) que nos permite deducir reglas de clasificación cualitativas aplicables a múltiples variantes (otros resultados experimentales de interés se muestran en el anexo II).

### **BRCA1 c.75C>T (p. Pro25=)**

Variante sinónima localizada en el exón 2, anotada como Clase-3 (clasificación clínica previa al estudio de *splicing*) en *BRCA1*Share. No hemos identificado en la literatura científica descripción de ningún estudio previo de *splicing* para esta variante.

Para caracterizar el posible efecto de la variante a nivel de *splicing*, se realizó una RT-PCR (cebadores localizados en los exones 2 y 6) a partir de RNA obtenido de sangre periférica de un portador y controles sanos. Los productos obtenidos se analizaron por electroforesis capilar (Figura 13A). No se observaron diferencias cualitativas. Tanto en el portador de la variante como en controles se detectó un producto principal que corresponde al transcrito de referencia, y 6 productos adicionales de muy baja intensidad que corresponden a diversos eventos de *splicing* alternativo individuales ( $\Delta 5$ ,  $\Delta 5q_{22}$ ,  $\Delta 3$ , y  $\nabla 4$ ), descritos previamente en controles, (Colombo et al, 2014), o a diversas combinaciones de los eventos de *splicing* anteriores ( $\Delta 3 + \Delta 5$  y  $\Delta 3 + \nabla 4 + \Delta 5$ ). Tampoco se observaron diferencias cuantitativas (contribución relativa de cada uno de los cuatro transcritos a la expresión global) relevantes. La secuenciación de la variante a partir de los productos de la RT-PCR (Figura 13B) demostró expresión bialélica balanceada en el portador de la variante. Este último dato es esencial, pues nos permite afirmar que la electroforesis capilar mostrada en la Figura 13A está capturando la expresión proveniente del alelo de la variante.

En conjunto, estos datos descartan que la variante afecte al *splicing* normal del gen. En nuestra opinión, este resultado negativo reduce significativamente la probabilidad de que la variante sinónima *BRCA1* c.75C>T (p. Pro25=) sea patogénica (no se predice efecto funcional a nivel de proteína, y ahora sabemos que tampoco tiene efecto funcional a nivel de RNA) por lo que, de acuerdo a las reglas de clasificación de ENIGMA, consideramos adecuado reclasificarla como Clase-1 (Clasificación clínica posterior al estudio de *splicing*).



**Figura 13. Estudio *in vitro* de *splicing* de la variante de *BRCA1* c.75C>T (p.Pro25=).** a) Análisis del *splicing* mediante RT-PCR/EC E2-E6. Se detectan los mismos transcritos Δ5 (220nt), Δ5q<sub>22</sub> (276nt), Δ3 (254nt), ▼4 (414nt), Δ3+Δ5 (166nt), Δ3+▼4+Δ5 (283nt), y transcrito de referencia (FL, 298nt) en controles sanos y en el portador de la variante c.75C>T (p.Pro25=). b) **Secuenciación Sanger.** El producto RT-PCR obtenido en un segundo ensayo (con cebadores E2 E7) se secuenció con un cebador E2 *sense*, donde se observa expresión bi-alélica.

### **BRCA1 c.594-2A>C (IVS9-2A>C)**

Variante genética localizada en el dinucleótido consenso del aceptor de *splicing* del exón 10. Durante mucho tiempo la variante ha estado anotada en las bases de datos internacionales como Clase-5, debido a estar localizada en un sitio consenso de *splicing* y causar la pérdida del exón 10 (Tesoriero et al 2005), una alteración que desplaza el marco de lectura e introduce un codón de parada prematuro. En la actualidad, la variante aparece anotada como Clase-1 tanto BRCAShare como ClinVar, reflejando el resultado de un estudio reciente llevado a cabo por los consorcios internacionales ENIGMA, CIMBA y BCAC en el que hemos participado (de la Hoya et al, 2016). Dicho estudio combina un análisis multifactorial que demuestra que la variante (para ser precisos, el haplotipo c.594-2A>C: 641A>G) es Clase-1, con un estudio de *splicing* (coordinado por nuestro grupo de investigación) en el que nos hemos propuesto caracterizar el efecto de la variante c.594-2A>C: 641A>G a nivel de *splicing* con el máximo detalle posible, asumiendo que dicho efecto, en cualquier caso, debería ser compatible con el estatus de variante no patogénica.

Nosotros aquí presentamos exclusivamente los datos de *splicing* realizados en nuestro laboratorio mediante RT-PCR/EC, secuenciación Sanger y PCR digital. Los resultados obtenidos por otros grupos del consorcio con otras técnicas (desde RNAseq hasta estudios en minigenes) confirman lo que aquí se va a exponer (de la Hoya et al, 2016).

El estudio de esta variante se ha llevado a cabo en una línea de células linfoblastoides (LCLs) portadora de la variante c.594-2A>C (Carrier 1) y los correspondientes controles no portadores. Los cultivos celulares de los que hemos extraído el RNA (tratados y sin tratar con puromicina) han sido suministrados por uno de los grupos de investigación integrantes del consorcio.

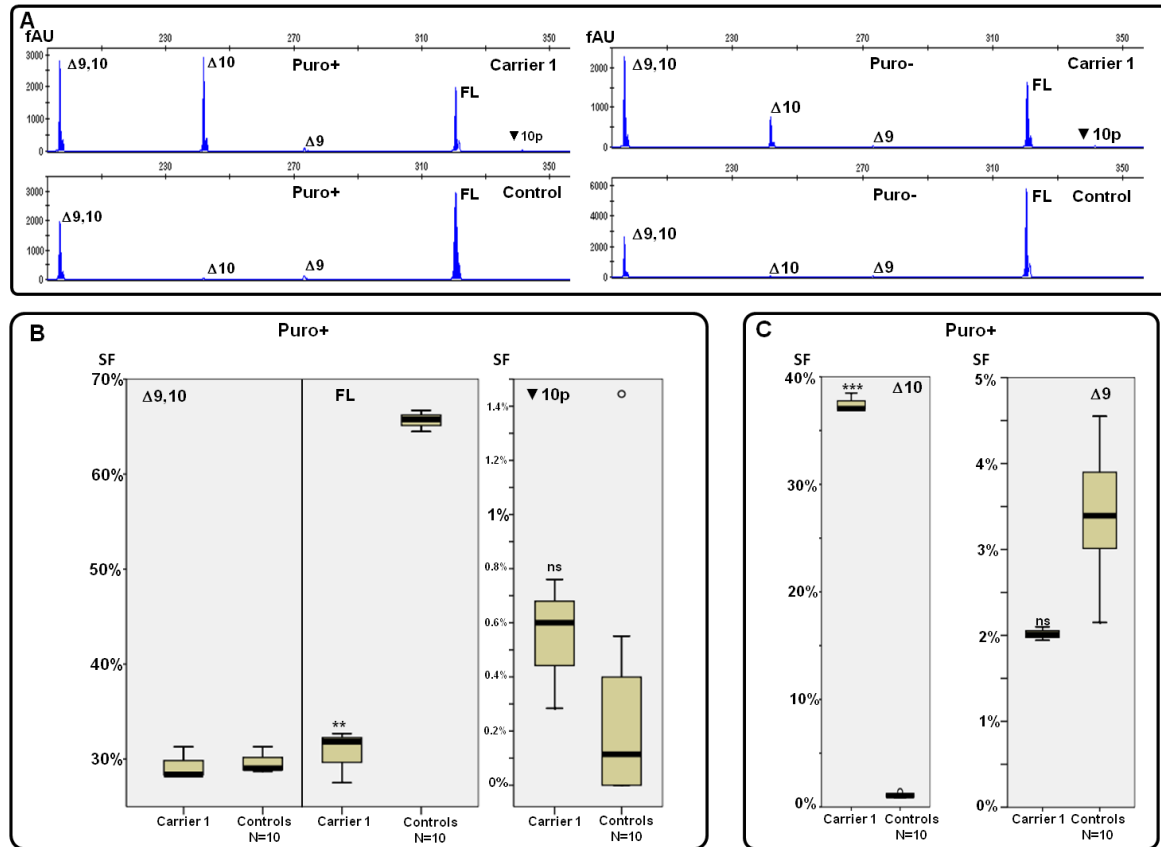
Para encontrar una explicación biológica plausible para justificar que no se incrementa el riesgo de cáncer de mama y ovario en portadores de la variante de *BRCA1* c.[594-2A>C: 641A>G], se realizó en primer lugar una caracterización del *splicing* alternativo que se produce en las regiones adyacentes al exón 10. Para dicha caracterización, realizamos análisis por electroforesis capilar a partir de RNAs extraídos de las LCLs (ver métodos). Se realizaron RT-PCRs (cebadores localizados entre los exones 8 y 11) a partir de RNA de LCLs portadoras de la variante c.594-2A>C y, y no portadoras. Los productos obtenidos se analizaron por electroforesis capilar detectándose 5 eventos de *splicing* alternativo en ambos tipos de muestra (Figura 14), incluyen 3 transcritos que mantienen la fase de lectura (el transcrito de referencia;  $\Delta 9,10$ ; y  $\nabla 10p^{21}$ ) y 2 transcritos que no la mantienen ( $\Delta 9$  y  $\Delta 10$ ). Todos los eventos,

salvo  $\nabla 10p^{21}$  (r.594-21\_594-1ins) han sido descritos previamente como parte del proceso normal de *splicing* alternativo en controles (Colombo et al, 2014). No se encontraron eventos específicos de la variante c.594-2A>C. En general, las variantes tratadas con puromicina (Puro + experimentos en Figura 14) indican que la expresión relativa de  $\Delta 9$ , 10 en portadores y controles es similar ( $\approx 29\%$ ), mientras la expresión relativa de  $\Delta 10$  aumenta ( $\approx 38\%$  vs.  $\approx 1\%$ ) y la del transcrito de referencia se reduce ( $\approx 31\%$  vs.  $\approx 66\%$ ). Las isoformas  $\Delta 9$  ( $<3\%$ ) y  $\nabla 10p^{21}$  ( $<1\%$ ) se expresaban como eventos minoritarios en todos los tipos de muestras analizadas (incluyendo controles), lo que demuestra que  $\nabla 10p^{21}$ , no descrito previamente, debe considerarse una nueva isoforma de *BRCA1* por *splicing* alternativo (bien es cierto que es una isoforma muy minoritaria).

Los datos proporcionados por los estudios de RT/PCR-EC no proporcionan una explicación obvia al hecho de que los portadores de la variante c.[594-2A>C;641A>G] no desarrollan características propias de una variante patogénica de *BRCA1*. Sin embargo, la ausencia de transcritos específicos del portador nos hizo proponer la hipótesis de que son los niveles de expresión de eventos que mantienen la fase de lectura (en particular  $\Delta 9,10$ ) los que explican la ausencia de riesgo asociado al haplotipo c.[594-2A>C; 641A>G].

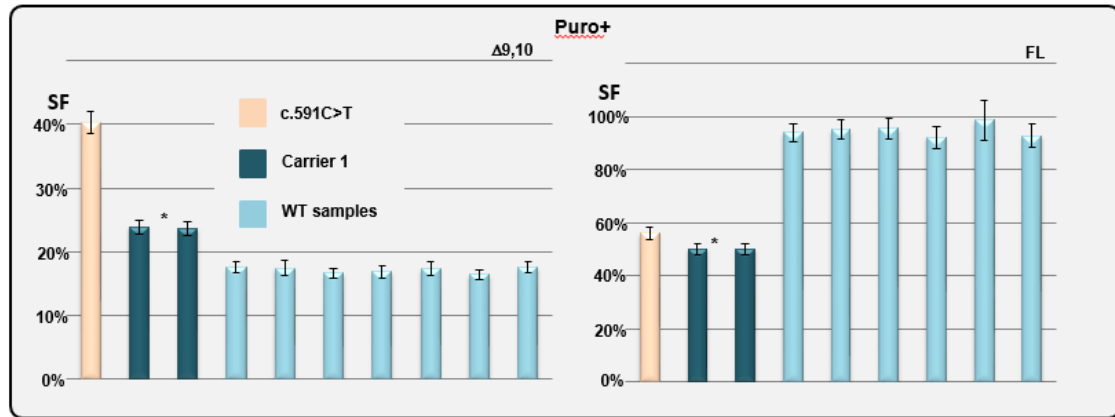
Hay que tener en cuenta, que en el mejor de los casos la electroforesis capilar es una técnica semi-cuantitativa, por lo que se decidió realizar análisis cuantitativos mediante PCR digital (dPCR). Estos estudios (Figura 14) revelan un incremento modesto, aunque significativo de  $\Delta 9,10$  en la muestra Carrier 1 ( $24\% \pm 0.9$ ) si se compara con controles (media del 17%), junto con una reducción del 50% de mensajeros de referencia. Esta reducción del 50% sugiere que el alelo de la variante no es capaz de producir mensajeros de referencia, extremo confirmado por otros grupos de investigación del consorcio ENIGMA mediante secuenciación Sanger de SNPs exónicos y mediante análisis en minigenes.

Los diversos estudios de *splicing* nos permiten concluir que el alelo c.594-2A>C no produce transcritos de referencia (en su lugar, produce niveles equivalentes de mensajeros PTC-NMD  $\Delta 10$ ). La única forma de reconciliar este dato con el análisis multifactorial (Clase-1) es asumir que los niveles de  $\Delta 9,10$  producidos por el alelo de la variante (esencialmente equivalentes a los observados en controles, probablemente algo superiores) deben ser suficientes para conferir haplosuficiencia.



**Figura 14. Análisis mediante RT-PCR/EC del *splicing* alternativo en el entorno de los exones 9 y 10 en una LCL portadora de la variante *BRCA1* c.[594-2A>C; 641A>G] (Carrier 1) y 10 LCLs controles.** La imagen A muestra ejemplos representativos de ensayos RT-PCR/EC E8.1-E11p realizados en LCLs tratadas y sin tratar con Puromicina. La intensidad de fluorescencia de cada pico (eje-Y) se expresa en unidades arbitrarias (fAU). En el análisis se detectaron el mensajero de referencia (FL), y cuatro isoformas de *splicing* alternativo, dos isoformas que mantienen la fase de lectura ( $\Delta 9$ , 10 y  $\nabla 10p^{21}$ ), y dos eventos que no mantienen la fase de lectura ( $\Delta 9$  y  $\Delta 10$ ). En este ejemplo, los transcritos de la isoforma  $\nabla 10p^{21}$  sólo se detectan en la muestra Carrier 1, pero ésta isoforma también se ha detectado anteriormente en controles, como se resume en la figura B. La presencia de la isoforma  $\nabla 10p^{21}$  en controles también ha sido confirmada por otros colaboradores del consorcio mediante RNAseq. Los diagramas de cajas de la Figura B representan la fracción de *splicing* (SF) de transcritos que mantienen la fase de lectura ( $\Delta 9$ , 10, FL y  $\nabla 10p$ ) detectados en el Carrier 1 (3 réplicas técnicas) y 10 controles. SF se expresa como el porcentaje que representa el área del pico de interés respecto al  $\Sigma$  del área de todos los picos anotados en el ensayo RT/PCR-EC. Este experimento se realizó utilizando el ensayo capilar E8.2-E11q.2. La fracción de mensajero de referencia es significativamente inferior en el Carrier 1 que en las muestras control. De forma análoga, en la figura C se representa la SF de los eventos PTC-NMD  $\Delta 9$  y  $\Delta 10$ . Los diagramas de cajas muestran el valor mínimo, percentil 25%, mediana, percentil 75% y valor máximo. (\*\* representan  $P \leq 0.01$ ) (\*\*\*) representan  $P \leq 0.001$ ) (ns= no es significativo). Figura adaptada de una publicación (de la Hoya et al, 2016).





**Figura 15. Contribución de  $\Delta 9,10$  y  $\nabla 9,10$  (FL) a la expresión global de *BRCA1* determinada por PCR digital en un portador de la variante *BRCA1* c.[594-2A>C; 641A>G] (Carrier 1) y muestras control.** Los análisis se realizaron en LCLs tratadas con Puromicina (puro +). Se utilizó un ensayo TaqMan FAM que reconoce la unión de los exones 8 y 11 ( $\Delta 9,10$ ), un ensayo TaqMan que reconoce la unión de los exones 9 y 10 ( $\nabla 9,10$ ), y un ensayo TaqMan VIC que reconoce la unión de los exones 23 y 24 (consideramos que captura la expresión global de *BRCA1*). La gráfica representa la proporción FAM/VIC, que corresponde a la contribución de  $\Delta 9,10$  (panel izquierdo) y  $\nabla 9,10$  (panel derecho) a la expresión global de *BRCA1*, o *splicing fraction* (SF). Se muestra el resultado de 2 réplicas técnicas para el Carrier 1, un control positivo (*BRCA1* c.591C>T incrementa los niveles de  $\Delta 9,10$ ), y 7 controles negativos. La precisión de cada medida se determinó con el software QuantStudio 3D Analysis Cloud. La SF de  $\Delta 9,10$  es superior en el Carrier 1 respecto a controles (24% en dos técnicas réplicas frente a un 17% en 7 muestras control, Mann-Whitney U test;  $p=0.028$  para la diferencia entre grupos). Por contra, la SF de  $\nabla 9,10$  (FL) se reduce casi un 50% (50% en dos técnicas réplicas frente a un 94% en 6 muestras control, Mann-Whitney U test;  $p=0.036$  para la diferencia entre grupos). Figura adaptada de la Hoya et al., 2016.

### **BRCA1 c.5152+4A>T (IVS18+4A>T)**

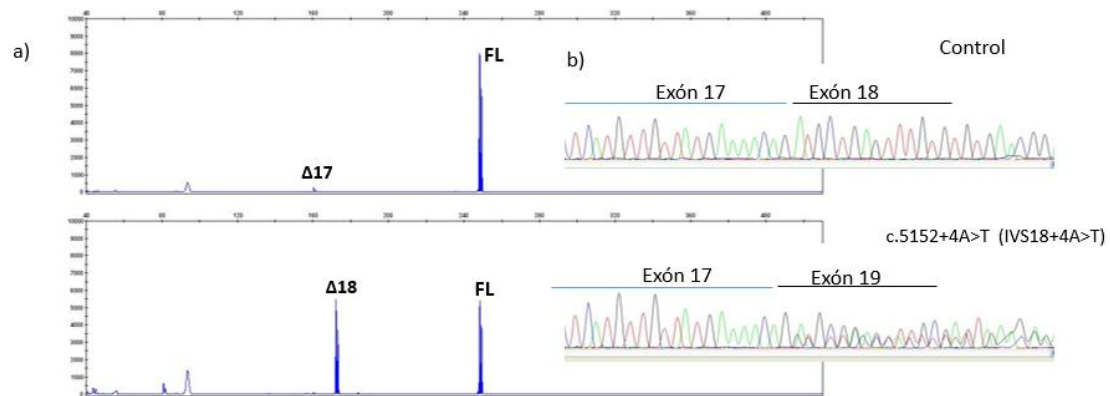
Variante intrónica rara localizada en el donador de *splicing* del exón 18. En el momento de redactar estas líneas no está descrita en *BRCA1Share*, ClinVar, o dbSNP. De acuerdo con los criterios del consorcio ENIGMA, se debe considerar Clase-3 (clasificación clínica previa al estudio de *splicing*). No hemos identificado en la literatura científica descripción de ningún estudio previo de *splicing* para esta variante.

Para la caracterización de esta variante se ha realizado una RT-PCR (cebadores localizados en los exones 16 y 19) a partir de RNA de sangre periférica obtenido de un portador y controles sanos. Los productos han sido analizados por electroforesis capilar (Figura 16a). El electroferograma del portador de la variante muestra diferencias cualitativas y cuantitativas: además del transcrito de referencia (FL) representado por el pico de 248nt, se detecta otro producto de intensidad similar de 172nt, que por tamaño corresponde muy aproximadamente a  $\Delta 18$  (el exón 18 de *BRCA1* tiene un tamaño de 78nt). Además, en el electroferograma de la muestra control se observa otro pico de poca intensidad que corresponde por tamaño (162nt) a  $\Delta 17$  (descrito previamente como evento de *splicing* alternativo minoritario por Colombo et al., 2014). La secuenciación (Figura 16b) de los productos de la RT-PCR, confirman la identidad de  $\Delta 18$ .

El exón 18 de *BRCA1* mantiene la fase de lectura. Sin embargo, codifica parte del dominio BRCT (esencial para la función de *BRCA1*), incluyendo algunos residuos críticos como la Glicina 1706 y la Alanina 1708 (se ha demostrado mediante análisis multifactorial que ciertas variantes *missense* en estas posiciones son inequívocamente patogénicas). Por ello, no tenemos dudas de que los transcritos  $\Delta 18$  deben considerarse deletéreos (de acuerdo con las reglas de clasificación de ENIGMA, una pérdida de nucleótidos que mantenga la fase de lectura eliminando algún residuo clínicamente relevante debe considerarse deletérea). Otras variantes que provocan igualmente la pérdida del exón 18 (5152+1G>C, 5152+1G>T, o 5152+5G>A) están consideradas patogénicas en BRCAShare.

Al no existir en el DNA del portador SNPs exónicos informativos, no hemos podido descartar formalmente la expresión de transcritos de referencia (transcritos que contengan el exón 18) a partir del alelo c.5152+4T. Siendo estrictos, esta limitación del estudio nos obligaría a considerar la variante Clase-4. Sin embargo, dada la proporción de transcritos  $\Delta 18$  que se observa tanto en la electroforesis capilar como en la secuencia Sanger (alrededor del 50%), consideramos probado que el alelo c.5152+4T no expresa exón 18 (o expresa una proporción

insignificante), por lo que consideramos razonable clasificarla como Clase-5 (Clasificación clínica posterior al estudio de *splicing*).



**Figura 16. Estudio *in vitro* de la variante de *BRCA1* c.5152+4A>T. a) Análisis de *splicing* mediante RT-PCR/EC E16-E19.** Se detecta el transcrito de referencia (FL, 248nt) en ambas muestras, Δ17 (162nt), muy minoritario, y Δ18 (172nt), mayoritario, en el portador de la variante c.5152+4A>T. **b) Secuenciación Sanger.** La secuenciación del producto de RT-PCR (cebador E16 *sense*) confirma la identidad de Δ18.

Exón/Intrón	HGVs <sup>1</sup>	Tipo	dbSNP	Clasificación inicial <sup>2</sup>	Predicción <i>In Silico</i> <sup>3</sup>	RT-PCR/EC <sup>4</sup>	Sanger <sup>4</sup>	dPCR <sup>4</sup>	Objetivo (Resultado) <sup>5</sup>
E1	c.75C>T (p.Pro25=)	Sinónima	rs80356839	Clase 3	No afecta	Patrón Normal	Expresión Bialélica	-	Reclasificar (Clase 1)
I2	c.81-13C>G (IVS2-13 C>G)	Intrónica	rs56328013	Clase 1 (MF)	Activación aceptor críptico	Patrón Normal	Expresión Bialélica	-	Caracterizar Alteración <i>Splicing</i> Tolerada (No informativo)
E6	c.230C>T (p.Thr77Met)	missense	rs80357209	Clase 3	Nuevo ESS/Pérdida ESE	Patrón Normal	Expresión Bialélica	-	Reclasificar (No informativo)
E8	c.528 G>A (p.Thr176=)	Sinónima	rs34545365	Clase 3	No afecta	Incremento $\Delta 9,10$	Expresión Bialélica	$\Delta 9,10$ (28%)	Reclasificar (Clase 1)
E9	c. 570C>T (p.Thr190=)	Sinónima	rs201536070	Clase 3	No afecta	Incremento $\Delta 9,10$	Expresión Bialélica	$\Delta 9,10$ (38%)	Reclasificar (Clase 1)
E9	c. 591C>T (p.Cys197=)	Sinónima	rs1799965	Clase 1 (MF)	Activación donador críptico	Incremento $\Delta 9,10$ <sup>#</sup>	Expresión Bialélica <sup>#</sup>	$\Delta 9,10$ (50%)	Caracterizar Alteración <i>Splicing</i> Tolerada (Informativo)
I9	c.594-2A>C (IVS9-2A>C)	Intrónica	rs80358033	Clase 1 (MF)	Pérdida Aceptor Nativo	Incremento $\Delta 9,10$	Expresión Bialélica	$\Delta 9,10$ (24%)	Caracterizar Alteración <i>Splicing</i> Tolerada (Informativo)
E11	c.773C>G (p.Pro258Arg)	Missense	rs80357225	Clase 3	No afecta	Patrón Normal	Expresión Bialélica	$\Delta 9,10$ (28%)	Reclasificar (No informativo)
E11	c.946A>G (p.Ser316Gly)	Missense	rs55874646	Clase 3	Nuevo ESS/Altera ESE	Incremento $\Delta 9,10$	Expresión Bialélica	$\Delta 9,10$ (50%)	Reclasificar (No informativo)
E11	c.1242C>T (p.Asp414=)	Sinónima	rs372400428	Clase 2	Nuevo ESS/Altera ESE	Incremento $\Delta 9,10$	Expresión Bialélica	$\Delta 9,10$ (30%)	Reclasificar (Clase1)
E12	c.4113G>A (p.Gln1371=)	Sinónima	rs147448807	Clase 2	No afecta	Patrón Normal	Expresión Bialélica	-	Reclasificar (Clase 1)
I14	c.4484+1G>T (IVS14+1G>T)	Intrónica	No descrita	Clase 4	Pérdida Donador Nativo	$\Delta 14$	$\Delta 14$ confirmado	-	Reclasificar (Clase 5)
I15	c.4675+11A>G (IVS15+11A>G)	Intrónica	No descrita	Clase 3	No afecta	Patrón Normal	No informativo <sup>##</sup>	-	Reclasificar (No informativa)
I15	c.4675+24G>A (IVS15+24G>A)	Intrónica	No descrita	Clase 3	No afecta	Patrón Normal	Expresión Bialélica	-	Reclasificar (Clase 1)
I18	c.5152+4A>T (IVS18+4A>T)	Intrónica	No descrita	Clase 3	Pérdida Donador Nativo	$\Delta 18$	$\Delta 18$ confirmado	-	Reclasificar (Clase 4)
E19	c.5175A>G (p.Glu1725=)	Sinónima	rs191373374	Clase 2	Nuevo ESS/Altera ESE	Patrón Normal	Expresión Bialélica	-	Reclasificar (Clase 1)
E19	c.5176A>G (p.Arg1726Gly)	Missense	rs80357501	Clase 3	Nuevo ESS/Altera ESE	Patrón Normal	Expresión Bialélica	-	Reclasificar (No informativa)
E20	c.5274A>G (p.Arg1758=)	Sinónima	rs758739620	Clase 3	Nuevo ESS/Altera ESE	Patrón Normal	Expresión Bialélica	-	Reclasificar (Clase 1)
I21	c.5333-2A>G (IVS21-2A>G)	Intrónica	No descrita	Clase 4	Pérdida Aceptor Nativo	$\Delta 22$ y $\Delta 22,23$	$\Delta 22$ confirmado	-	Reclasificar (Clase 5)
E21	c.5357T>C (p.Leu1786Pro)	Missense	rs398122697	Clase 3	Altera ESE	Patrón Normal	Expresión Bialélica	-	Reclasificar (No informativo)
E23	c.5412C>T (p.Val1823=)	Sinónima	rs730881456	Clase 3	Altera ESE	Patrón Normal	Expresión Bialélica	-	Reclasificar (Clase 1)

**Tabla 8. Estudios de *Splicing* en portadores de variantes genéticas en *BRCA1*.** <sup>1</sup> Se siguen las reglas publicadas por la *Human Genome Variation Society* (<http://vRNAomen.hgvs.org/>), utilizando como referencia el transcrito ENST00000357654. <sup>2</sup> Para la clasificación previa a nuestro estudio de *splicing* se consultaron las bases de datos BRCAShare y ClinVar, y se siguieron los criterios de clasificación cualitativa establecidos por el consorcio ENIGMA ([www.enigmaconsortium.com](http://www.enigmaconsortium.com)), MF (clasificación multifactorial). La tabla refleja la información disponible en las bases de datos en el momento de realizar los estudios. Las bases de datos BRCAShare y ClinVar se actualizan con frecuencia, por lo que una consulta en el momento actual puede dar otro resultado. Por ejemplo, el 29 de Junio de 2017 el consorcio ENIGMA volcó en ClinVar una actualización de sus propias reglas de clasificación según la cual, una variante sinónima con predicción *in silico* de *splicing* negativo se considera Clase-2. <sup>3</sup> Los estudios *in silico* de *splicing* se realizaron en la plataforma on-line Human *Splicing* Finder (<http://www.umd.be/HSF3/>). El posible efecto en donadores/aceptores de *splicing* se analizó con los algoritmos MaxEnt y HSF. En la tabla se muestra el resultado del algoritmo que con mayor probabilidad predice un efecto. En caso de no predecirse ningún efecto a nivel de donadores/aceptores, se muestra el posible efecto a nivel de secuencias reguladoras detectado con cualquiera de los algoritmos que con este propósito se incluyen en la plataforma HSF3. ESE (secuencia exónica que potencia el *splicing*). ESS (Secuencia exónica que inhibe el *splicing*).<sup>4</sup> Los datos más representativos se muestran en resultados y Anexo I. El % indicado en dPCR se refiere a la contribución de  $\Delta 9,10$  a la expresión global de *BRCA1*. <sup>5</sup> Se propone una clasificación clínica basada exclusivamente en la interpretación de los resultados de los estudios de *splicing*. Pueden existir datos genéticos y/o funcionales que justifiquen otra clasificación. <sup>#</sup> El análisis inicial de la variante c. 591C>T (p.Cys197=) mediante RT/PCR/EC y secuenciación Sanger se realizó hace años (Dosil et al, 2010). Aquí hemos realizado análisis de RT/PCR confirmatorios y análisis por PCR digital. Recientemente se ha demostrado mediante análisis multifactorial que la variante no tiene relevancia clínica (de la Hoya et al, 2016). <sup>##</sup> La ausencia de SNPs exónicos nos ha impedido confirmar expresión Bialélica.

Exón/Intrón	HGVS <sup>1</sup>	Tipo	dbSNP	Clasificación inicial <sup>2</sup>	Predicción <i>In silico</i> <sup>3</sup>	RT-PCR <sup>4</sup>	Sanger <sup>4</sup>	dPCR <sup>4</sup>	Objetivo (Resultado) <sup>5</sup>
I2	c.68-7T>A (IVS2-7T>A)	Intrónica	rs81002830	Clase-1 (MF)	No afecta	Incremento Δ3	-	Incremento Δ3	Caracterizar Alteración <i>Splicing</i> Tolerada (Informativo)
E3	c.223G>C (p.Ala75Pro)	Missense	rs28897701	Clase-1(MF)	Nuevo ESS/Altera ESE	Patrón Normal	Expresión Bialélica	Δ3 normal	Caracterizar Alteración <i>Splicing</i> Tolerada (No Informativo)
E3	c.280C>T (p.Pro94Ser)	Missense	rs80358531	Clase-3	Nuevo ESS/Altera ESE	Patrón Normal	Expresión Bialélica	Δ3 normal	Reclasificar (No Informativo)
E7	c.526A>T (p.Thr176Ser)	Missense	rs551319375	Clase-3	Altera ESE	Patrón Normal	Expresión Bialélica	-	Reclasificar (No informativo)
E9	c.740T>C (p.Ile247Thr)	Missense	rs80358962	Clase-3	No afecta	Patrón Normal	Expresión Bialélica	-	Reclasificar (No informativo)
E11	c.1518C>T (p.Phe206=)	Sinónima	No descrita	Clase-3	Nuevo ESS/Altera ESE	Patrón Normal	Expresión Bialélica	-	Reclasificar (Clase-1)
E11	c.6246A>G (p.Glu2082=)	Sinónima	No descrita	Clase-3	Activación aceptor críptico Activación donador críptico	Patrón Normal	Expresión Bialélica	-	Reclasificar (Clase-1)
E11	c.6742C>G (p.His2248Asp)	Missense	No descrita	Clase-3	Nuevo ESS/Altera ESE	Patrón Normal	Expresión Bialélica	-	Reclasificar (No informativo)
E13	c.7007G>A (p.Arg2336His)	Missense	rs28897743	Clase-5	Pérdida Donador Nativo	Δ12,13 y Δ13 <sup>#</sup>	Δ12,13 y Δ13 confirmados	-	Estudios previos discrepantes (Informativo)
E14	c.7283T>G (p. Leu2428Trp)	Missense	No descrita	Clase-3	Nuevo ESS/Altera ESE	Patrón Normal	Expresión Bialelica	-	Reclasificar (No informativa)
I16	c.7805+17 G>A (IVS16+17G>A)	Intrónica	rs371300428	Clase-3	No afecta	Patrón Normal <sup>#</sup>	Expresión Bialélica	-	Reclasificar (Clase-1)
I16	c.7805+23G>A (IVS16+23G>A)	Intrónica	rs113653200	Clase-3	No afecta	Patrón Normal <sup>#</sup>	Expresión Bialélica	-	Reclasificar (Clase-1)
I19	c.8488-1G>A (IVS19-1G>A)	Intrónica	rs397507404	Clase-4	Pérdida Aceptor Nativo	Δ20p <sub>12</sub>	Δ20p <sub>12</sub> confirmado	-	Reclasificar (Clase-3) <sup>###</sup>
E20	c.8499G>A (p.Lys2833=)	Sinónima	rs558819788	Clase-2	Altera ESE	Patrón Normal	Expresión Bialélica	-	Reclasificar (Clase-1)
E20	c.8592C>T (p.Ala2864=)	Sinónima	rs32945197	Clase-3	No afecta	Patrón Normal	Expresión Bialélica	-	Reclasificar (Clase-1)
E22	c.8767G>A (p.Glu2923Lys)	Missense	No descrita	Clase 3	Altera ESE	Patrón Normal	Expresión Bialélica	-	Reclasificar (No informativo)
E23	c.9086C>T (p.Ala3029Val)	Missense	rs80359162	Clase 3	Activación donador críptico	Patrón Normal <sup>#</sup>	Expresión Bialélica	-	Reclasificar (No informativo)

**Tabla 9. Estudios de *Splicing* en portadores de variantes genéticas en *BRCA2*.** <sup>1</sup> Se siguen las reglas publicadas por la *Human Genome Variation Society* (<http://vRNAomen.hgvs.org/>), utilizando como referencia el transcrito ENST00000380152. <sup>2</sup>Para la clasificación previa al estudio de *splicing* se consultaron las bases de datos BRCAShare y ClinVar, y se siguieron los criterios de clasificación cualitativa establecidos por el consorcio ENIGMA ([www.enigmaconsortium.com](http://www.enigmaconsortium.com)), MF (clasificación multifactorial). La tabla refleja la información disponible en las bases de datos en el momento de realizar los estudios. Las bases de datos BRCAShare y ClinVar se actualizan con frecuencia, por lo que una consulta en el momento actual puede dar otro resultado. Por ejemplo, el 29 de Junio de 2017 el consorcio ENIGMA volcó en ClinVar una actualización de sus propias reglas de clasificación según la cual una variante sinónima con predicción *in silico* de *splicing* negativo se considera Clase-2. <sup>3</sup>Los estudios *in silico* de *splicing* se realizaron en la plataforma on-line *Human Splicing Finder* (<http://www.umd.be/HSF3/>). El posible efecto en donadores/aceptores de *splicing* se analizó con los algoritmos MaxEnt y HSF. En la tabla se muestra el resultado del algoritmo que con mayor probabilidad predice un efecto. En caso de no predecirse ningún efecto a nivel de donadores/aceptores, se muestra el posible efecto a nivel de secuencias reguladoras detectado con cualquiera de los algoritmos que con este propósito se incluyen en la plataforma HSF3. ESE (secuencia exónica que potencia el *splicing*). ESS (Secuencia exónica que inhibe el *splicing*). <sup>4</sup> Los datos más representativos se muestran en resultados y Anexo I. El % indicado en dPCR se refiere a la contribución de Δ3 a la expresión global de *BRCA2*. <sup>5</sup>Se propone una clasificación clínica basada exclusivamente en la interpretación de los resultados de los estudios de *splicing*. Pueden existir datos genéticos y/o funcionales que justifiquen otra clasificación. <sup>#</sup> RT-PCR/Agarosa. <sup>###</sup>Ejemplo de variante donde el estudio de *splicing* es informativo, pero no reduce la incertidumbre en torno a la clasificación de la variante, sino que la aumenta. Esto se debe a que el estudio de *splicing* muestra que la variante genera transcritos que potencialmente podrían codificar una proteína funcional.

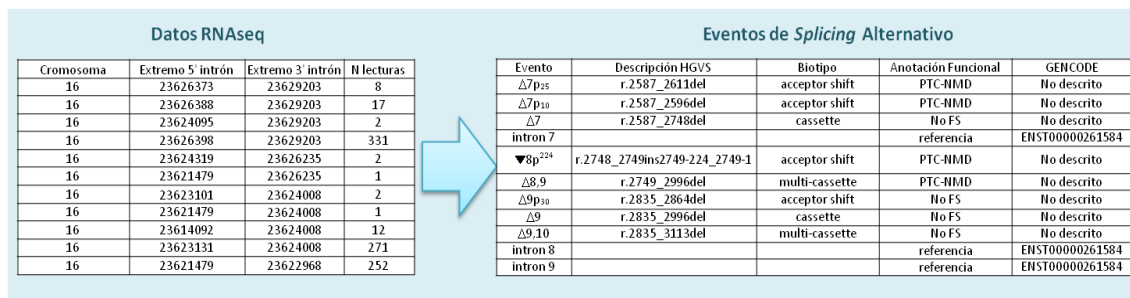
## II. Caracterización del *splicing* alternativo en *RAD51C*, *RAD51D* y *PALB2*.

### 1. Anotación de eventos

Para la caracterización del *splicing* alternativo se partió de dos fuentes de datos primarias independientes:

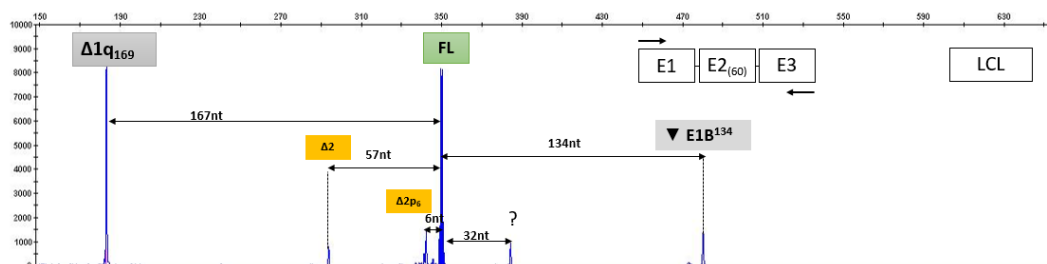
- 1) Datos de RNAseq generados a partir de LCLs (tratadas y sin tratar con un inhibidor del mecanismo NMD), tejido normal de mama, y tejido de fimbria ovárica. En esencia, los datos de RNAseq contenían el resultado final de alinear las secuencias contra un genoma de referencia mediante los alineadores STAR y TopHat (estos alineadores producen parejas de coordenadas genómicas que definen potenciales intrones y, por tanto, potenciales eventos de *splicing*. Cada pareja de coordenadas genómicas va asociada a un número de lecturas concretas que puede contener información cuantitativa). Los datos de RNAseq fueron proporcionados por dos laboratorios colaboradores en el marco del consorcio ENIGMA.
- 2) El análisis realizado en nuestro laboratorio por electroforesis capilar (y en algunos casos también secuenciación Sanger) de los productos generados en múltiples reacciones de RT-PCR realizadas sobre los tres *loci* de interés (ensayos RT-PCR/EC). Todos los ensayos han sido realizados a partir de RNAs purificados de cuatro tipos de muestras biológicas: sangre periférica, LCLs, tejidos comerciales de mama y tejidos comerciales de ovario.

A partir de las dos fuentes de datos realizamos una anotación manual de potenciales eventos de *splicing* alternativo. Por anotación manual de los datos de RNAseq entendemos transformar las coordenadas genómicas en eventos de *splicing* alternativo descritos en notación HGVS respecto a un mensajero de referencia, tal y como se describe en la Figura 17.



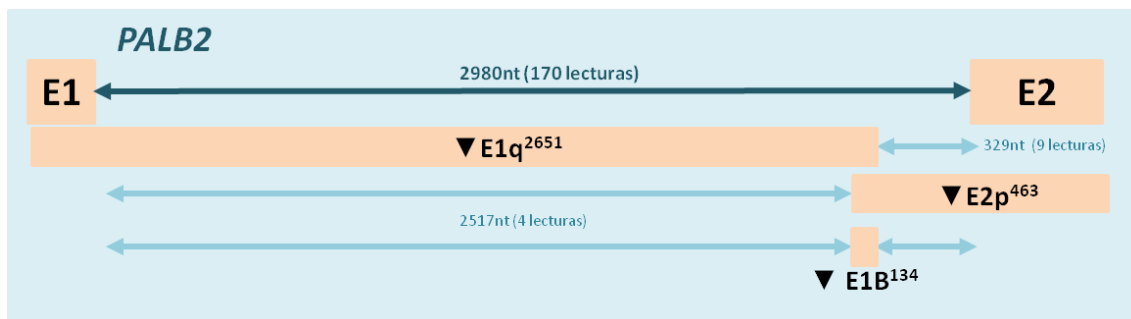
**Figura 17. Ejemplo de anotación manual de los datos de RNAseq.** En la tabla de la izquierda se muestra un ejemplo concreto correspondiente al locus *PALB2* del tipo de datos generados por los algoritmos de alineamiento STAR y TopHat. Se proporcionan las coordenadas genómicas (GRCh38) correspondientes a los huecos en el alineamiento (posibles intrones) y el número de lecturas correspondientes. A partir de esos datos, de forma manual definimos posibles eventos de *splicing* alternativo, su descripción en nomenclatura HGVS, su biotipo y anotación funcional (Mudge et al, 2013), y comprobamos si se han descrito en GENCODE transcritos compatibles con el evento en cuestión (*GENCODE comprehensive set*). PTC-NMD (Se predice un mensajero con un codón de parada prematuro y degradación selectiva vía mecanismo NMD), No FS (el evento de *splicing* mantiene el marco de lectura del mensajero de referencia), Referencia (lecturas correspondientes a un intrón presente en el mensajero de referencia de *PALB2*, en este caso el mensajero *PALB2* ENST00000261584).

Para realizar anotaciones de *splicing* a partir del análisis por RT-PCR/EC, nos fijamos en la diferencia de tamaño entre los distintos picos y el pico correspondiente al transcrito de referencia, y analizamos si esa diferencia de tamaño es compatible con el uso de un determinado par GT/AG presente en la secuencia genómica de referencia, tal y como se ilustra en la Figura 18.



**Figura 18. Ejemplo representativo de experimento RT-PCR/EC realizado entre los exones 1 y 3 de *PALB2*.** Este tipo de experimentos se realizaron en LCLs (N=15), Sangres (N=8), tejido de mama (N=8), y ovario (N=8). En el caso de mama y ovario la N se refiere a réplicas técnicas. En los 4 tejidos la RT-PCR genera un mínimo de 6 productos (picos) que pueden corresponder a otros tantos transcritos. Una vez identificado el pico correspondiente al mensajero de referencia (352nt), la naturaleza del resto de picos se infiere en base a las diferencias de tamaño y la presencia de secuencias canónicas GT/AG compatibles. Así, el pico de 343nt podría corresponder en principio a los eventos  $\Delta 1q_6$ ,  $\Delta 2p_6$ ,  $\Delta 2q_6$  o  $\Delta 3p_6$ , pero sólo  $\Delta 2p_6$  es compatible con un intrón GT/AG. En gris aparecen representados los eventos confirmados (detectados también por RNAseq), y en naranja los eventos no confirmados (detectados sólo por RT-PCR/EC). En verde se representa el pico correspondiente al transcrito de referencia. El símbolo de interrogación señala un pico para el cual no hemos sabido inferir ninguna anotación. Podría tratarse de un evento real determinado por secuencias no canónicas (Sibley et al, 2016). También podría corresponder a una combinación compleja de varios eventos de *splicing* en el mismo transcrito, o podría, por supuesto, ser un producto inespecífico de la PCR. En cualquier caso, es ejemplo de lo que hemos denominado **potenciales eventos de *splicing* no anotados**.

En el caso de potenciales nuevos exones, la combinación de datos de RNAseq y RT-PCR/EC resulta fundamental para eliminar ambigüedades experimentales, tal y como se ilustra en la Figura 19.



**Figura 19. La combinación de datos de RNAseq y RT-PCR/EC resulta fundamental para definir nuevos exones.** Tal y como se indica en la figura, los datos de RNAseq para el *locus PALB2* identificaron además del intrón 1 de referencia (2980nt; 170 lecturas), dos potenciales intrones de 2517nt (4 lecturas) y 329nt (9 lecturas). Los datos de RNAseq se pueden interpretar de dos formas alternativas: uso de un donador de *splicing* alternativo para el exón 1 (▼ E1q<sup>2651</sup>) más uso de un aceptor de *splicing* alternativo para el exón 2 (▼ E2p<sup>463</sup>), o bien inserción de un nuevo exón de 134nt entre los exones 1 y 2 de referencia (▼ E1B<sup>134</sup>). Por otro lado, los datos de RT/PCR-EC (Figura 44) detectan un posible transcrito 134 nt mayor que el transcrito de referencia para el que es muy difícil inferir una anotación basándose exclusivamente en el tamaño. Sin embargo, la existencia de un nuevo intrón de 134nt situado a 2517nt del exón 1 (y a 329nt del exón 2) parece evidente cuando se combinan las dos fuentes de datos.

Una vez realizadas las anotaciones manuales de los datos de RNAseq y RT-PCR/EC, comparamos los datos y definimos eventos de *splicing* alternativo agrupados en función de su grado de evidencia experimental de acuerdo con los siguientes criterios:

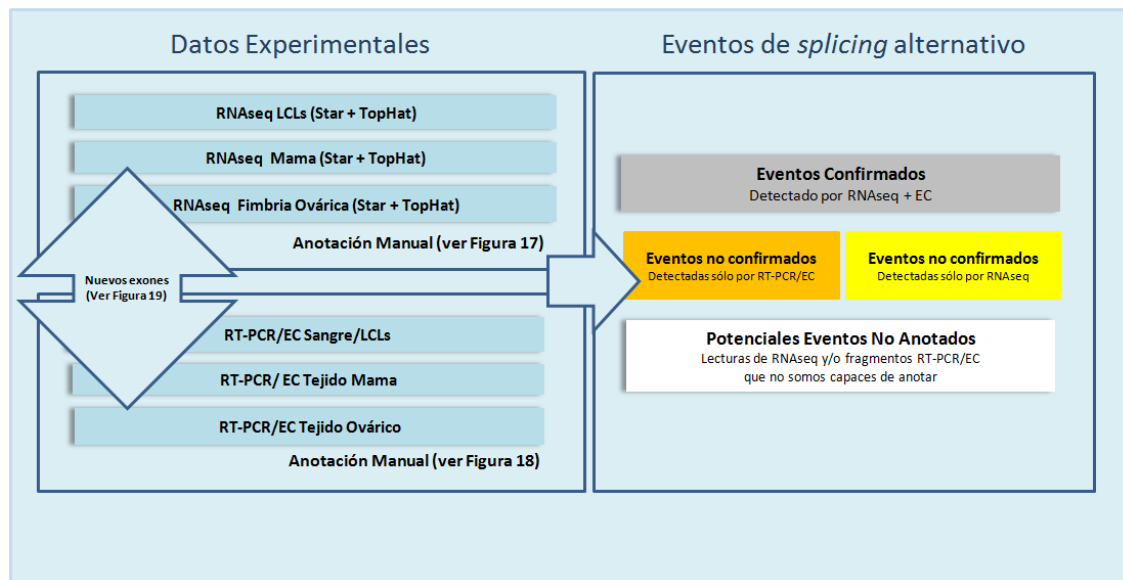
**Eventos confirmados** son eventos anotados a partir de datos de RNAseq (o en algunos pocos casos secuenciados por Sanger) y detección de picos de tamaño compatibles por electroforesis capilar.

**Eventos no confirmados** son eventos anotados a partir de datos de RNAseq (típicamente, muy pocas lecturas en RNAseq) pero sin detección de picos compatibles en electroforesis capilar, o eventos anotados (inferidos) a partir de picos de electroforesis capilar, pero sin lecturas que las confirmen por RNAseq (ni secuenciación Sanger).

**Potenciales eventos de *splicing* no anotados.** Lecturas de RNAseq y/o picos detectados por RT-PCR/EC que no tienen una interpretación obvia en término de *splicing* y que por tanto no hemos sido capaces de anotar. Probablemente, muchas de estas señales representarán artefactos metodológicos, aunque otras pueden representar eventos reales de expresión muy minoritaria.



Para distinguir de forma clara y visual a lo largo de esta tesis el grado de evidencia experimental que apoya cada uno de los eventos de *splicing* se ha utilizado el siguiente código de colores: gris para eventos confirmados, amarillo para eventos no confirmados por RT-PCR/EC, y naranja para eventos no confirmados por RNAseq. A continuación, se representa un esquema para facilitar la comprensión del proceso de trabajo que acabamos de describir (Figura 20).



**Figura 20. Esquema seguido para la caracterización de eventos de *splicing* alternativo en *RAD51C*, *RAD51D* y *PALB2*.** Los experimentos de RNAseq se realizaron a partir de 3 tipos de muestras biológicas, y se analizaron en paralelo con los algoritmos de alineamiento de secuencias TopHat y STAR. Los experimentos de RT-PCR/EC se realizaron a partir de 4 tipos de muestras biológicas. Los eventos detectados en las dos aproximaciones se compararon para generar una lista final de eventos de alta (**confirmados**) y baja confianza (**no confirmados**). Los eventos de alta confianza incluyen nuevos exones que se anotan combinando los datos de RNAseq y RT-PCR/EC. Tanto los datos de RNAseq como los datos de RT-PCR/EC generan señales que no hemos sido capaces de anotar y denominamos **potenciales eventos no anotados**.

A continuación, se muestra la anotación de eventos realizada para cada gen de susceptibilidad estudiado, algunos ejemplos representativos de los ensayos RT-PCR/EC realizados, y los resultados estratificados por tejido.

## 2. Caracterización del *splicing* alternativo en el gen *RAD51C*

En total se han podido confirmar 56 eventos de *splicing* alternativo en el gen *RAD51C* (Tabla 10), de los cuáles, 36 introducen un codón de parada prematuro (PTC-NMD), 18 mantienen la fase de lectura (No-Fs), uno desplaza el marco de lectura en el último exón (FS-alternative STOP) y uno elimina el ATG iniciador de la traducción (Non-coding). Otros eventos no confirmados se describen en la Tabla 11.

Evento <sup>1</sup>	Descripción HGVS <sup>2</sup>	GENCODE <sup>3</sup>	Biotipo <sup>4</sup>	Anotación Funcional <sup>4</sup>	Lecturas <sup>5</sup>	RT-PCR/EC
$\Delta 1q^{156}$	r.-11_145del	ENST00000487921.5	donor shift	Non-coding	28	PCR E1E2/PCR E1E3
$\Delta 1q_{156} + \nabla E1A^{300}$	r.-11_145del+r.145_146ins145+563_145+862	No descrito	mixed	PTC-NMD	11,5	PCR E1E2/PCR E1E3
$\Delta 1q_{103}$	r.43_145del	ENST00000461271.5	donor shift	PTC-NMD	97,5	PCR E1E2/PCR E1E3
$\Delta 1q_{103} + \nabla E1A^{300}$	r.43_145del+r.145_146ins145+563_145+862	No descrito	mixed	PTC-NMD	16,5	PCR E1E2/PCR E1E3
$\Delta 1q_{103} \Delta 2p_3$	r.43_148del	No descrito	mixed	PTC-NMD	3	PCR E1E2/PCR E1E3
$\Delta 1q_{103} \Delta 2$	r.43_404del	No descrito	mixed	PTC-NMD	1	PCR E1E3
$\nabla 1q_{171}$	r.145_146ins145+1_145+171	No descrito	donor shift	PTC-NMD <sup>^</sup>	9	Figura 21
$\nabla E1A^{351}$	r.145_146ins145+563_145+914	No descrito	casette	No FS	18	Figura 21
$\nabla E1A^{300}$	r.145_146ins145+563_145+862	No descrito	casette	No FS	29,5	Figura 21
$\nabla E1A^{461}$	r.145_146ins145+563_145+1023	No descrito	casette	PTC-NMD	29,25	PCR E1E3
$\nabla E1B^{56}$	r.145_146ins145+968_145+1023	No descrito	casette	PTC-NMD	19	Figura 21
$\nabla E1A^{300} + \nabla E1B^{56}$	r.145_146ins145+563_145+862+r.145_146ins145+968_145+1023	No descrito	multicasette	PTC-NMD	1,5	PCR E1int E2int*
$\nabla 2p^{28}$	r.145_146ins146-28_146-1	No descrito	acceptor shift	PTC-NMD	1	PCR E1int-E2int*
$\nabla E1A^{461} + \Delta 2p_3$	r.145_146ins145+563_145+1023+r.146_148del	No descrito	mixed	PTC-NMD	1	PCR E1int-E2int no se detecta
$\Delta 2p_3$	r.146_148del	No descrito	acceptor shift	No FS	1	Figura 21
$\Delta 2$	r.146_404del	ENST00000584617.5	casette	PTC-NMD	1	PCR E1E3
$\Delta 2q_{175}$	r.230_404del	No descrito	donor shift	PTC-NMD	1	PCR E1E3
$\Delta 2q_{22}$	r.383_404del	No descrito	donor shift	PTC-NMD	1	Figura 22
$\nabla 2q^{27}$	r.404_405ins404+1_404+27	No descrito	donor shift	PTC-NMD <sup>^</sup>	20	PCR E2E3/PCR E2E4

Evento <sup>1</sup>	Descripción HGVS <sup>2</sup>	GENCODE <sup>3</sup>	Biotipo <sup>4</sup>	Anotación Funcional <sup>4</sup>	Lecturas <sup>5</sup>	RT-PCR/EC
Δ3	r.405_571del	No descrito	casette	PTC-NMD	73	Figura 22
Δ3,4	r.405_705del	No descrito	multicasette	PTC-NMD	1	PCR E2E5
Δ3_8	r.405_1026del	No descrito	multicasette	PTC-NMD	1	PCR E2E9
Δ3q <sub>114</sub>	r.458_571del	No descrito	donor shift	No FS	1	Figura 22
Δ3q <sub>114</sub> _Δ4p <sub>81</sub>	r.458_652del	No descrito	mixed	No FS	38 <sup>#</sup>	Figura 22
Δ4p <sub>81</sub>	r.572_652del	No descrito	acceptor shift	No FS	10,5	Figura 22
Δ4p <sub>111</sub>	r.572_682del	No descrito	acceptor shift	No FS	1	PCR E3E5
Δ4	r.572_705del	No descrito	casette	PTC-NMD	15,5	PCR E3E5
Δ4,5	r.572_837del	No descrito	multicasette	PTC-NMD	1	PCR E3E8
Δ4+▼E4A <sup>145</sup>	r.572_705del+r.705_706ins705+2549_705+2693	No descrito	mixed	No FS	3	PCR E3E5
▼E4A <sup>145</sup>	r.705_706ins705+2549_705+2693	ENST00000487525.5	casette	PTC-NMD	235	Figura 23
▼E4B <sup>120</sup>	r.705_706ins705+3160_705+3279	ENST00000425173.5	casette	No FS	29,75	PCR E4E5*
▼E4C <sup>48</sup>	r.705_706ins705+4462_705+4509	No descrito	casette	No FS	1	PCR E4E5*
Δ5	r.706_837del	No descrito	casette	No FS	1	PCR E4E7
Δ5q <sub>15/14</sub> _7p <sub>34/35</sub> **	r.827_942del	No descrito	mixed	PTC-NMD	-	PCR E4E8
▼E5A <sup>57</sup>	r.837_838ins837+606_837+662	No descrito	casette	No FS	20	PCR E5E7/ E5E8
▼E5B <sup>75</sup>	r.837_838ins837+4016_837+4091	No descrito	casette	No FS	1	PCR E5E7/ E5E8
▼E5C <sup>100</sup>	r.837_838ins837+10526_837+10625	No descrito	casette	PTC-NMD	13,25 <sup>#</sup>	PCR E5E7/ E5E8
▼E5C <sup>48</sup>	r.837_838ins837+10577_837+10625	No descrito	casette	No Fs	1	PCR E5E7/ E5E8
▼E5A <sup>57</sup> +▼E5C <sup>100</sup>	r.837_838ins837+606_837+662+r.837_838ins837+10526_837+10625	No descrito	multi-casette	PTC-NMD	1	PCR E3E8
▼E5A <sup>57</sup> +▼E5C <sup>48</sup>	r.837_838ins837+606_837+662+r.837_838ins837+10577_837+10625	No descrito	multi-casette	No Fs	1	PCR E3E8
Δ6	r.838_904del	ENST00000475762.5	casette	PTC-NMD	7	Figura 24
Δ6,7	r.838_965del	No descrito	multicasette	PTC-NMD	1	Figura 24
Δ6,7 <sub>128</sub> +▼8p <sup>3</sup>	r.838_965del+r.965_966ins965+1592_965+1663	No descrito	mixed	PTC-NMD	2	PCR E5E7/E5E8
Δ6_8	r.838_1026del	No descrito	multicasette	No FS	4,5	PCR E5E9

Evento <sup>1</sup>	Descripción HGVS <sup>2</sup>	GENCODE <sup>3</sup>	Biotipo <sup>4</sup>	Anotación Funcional <sup>4</sup>	Lecturas <sup>5</sup>	RT-PCR/EC
▼7p <sup>22</sup>	r.904_905ins905-22_905-1	No descrito	acceptor shift	PTC-NMD	1	PCR E5E9
Δ7	r.905_965del	ENST00000584804.1	cassette	PTC-NMD	76,5 <sup>#</sup>	Figura 24
Δ7+▼8p <sup>3</sup>	r.905_r.965del+r.965_966ins966-3_966-1	No descrito	mixed	PTC-NMD	2,5	PCR E5E8
Δ7,8	r.905_1026del	ENST00000578151.1	multicassette	PTC-NMD	5	PCR E5E9
▼E7A <sup>72</sup>	r.965_966ins965+1592_965+1663	No descrito	cassette	PTC-NMD	14,5	PCR E5E9
▼E7A <sup>72</sup> +▼8p <sup>3</sup>	r.965_966ins965+1592_965+1663+r.965_966ins966-3_966-1	No descrito	mixed	PTC-NMD	12,25	PCR E5E9
▼E7B <sup>122</sup>	r.965_966ins966-2298_966-2177	No descrito	cassette	PTC-NMD	10,25 <sup>#</sup>	Figura 25
▼E7B <sup>122</sup> +▼8p <sup>3</sup>	r.96r.965_966ins966-2298_966-2177+r.965_966ins966-3_966-1	No descrito	mixed	PTC-NMD	9	PCR E5E8/E7E9*
▼8p <sup>3</sup>	r.965_966ins966-1_966-3	ENST00000413590.5	acceptor shift	No FS	102	Figura 25
Δ8	r.966_1026	No Descrito	cassette	PTC-NMD	5,5	Figura 25
▼9p <sup>31</sup>	r.1026_1027ins1027-31_1027-1	No Descrito	acceptor shift	FS-alternative STOP	30 <sup>#</sup>	PCR E7E9*
Δ9p <sub>6</sub>	r.1027_1032del	No Descrito	acceptor shift	No FS	1,2	Figura 25

**Tabla 10. Eventos de *splicing* alternativo en RAD51C confirmados.** <sup>1</sup> En Materiales y Métodos se detalla el sistema de designación de eventos. <sup>2</sup> Se siguen las reglas publicadas por la *Human Genome Variation Society* (<http://vrnaomen.hgvs.org/>). <sup>3</sup> Se utilizó Ensembl (<http://www.ensembl.org>) para comprobar si el evento aparece descrito en la compilación 26 de GENCODE. <sup>4</sup> Biotipo y Anotación Funcional según los criterios desarrollados por GENCODE (Mudge et al, 2013). <sup>5</sup> Se muestra el número de lecturas obtenido en el experimento realizado en LCL+puro (media de los valores obtenidos con STAR y TopHat). En el mismo experimento, el número de lecturas (media de STAR y TopHat) correspondiente a cada uno de los 8 intrones de referencia osciló entre 849 y 1654 (media 1249). La cobertura alcanzada en los experimentos realizados en otras muestras biológicas fue sensiblemente inferior, de tal modo que el valor medio de un intrón de referencia fue 291 en la muestra LCL-puro, 358 en Mama, y 52 en fimbria ovárica. <sup>#</sup> Se destacan eventos con número de lecturas muy discrepantes en STAR y TopHat (por ejemplo, Δ3q<sub>114</sub>\_Δ4p<sub>81</sub> tiene 2 lecturas en STAR y 74 en TopHat). \*Eventos que han sido detectados en los ensayos señalados, pero se detectan como productos de muy baja intensidad. Y sólo se detectan en algunas muestras. Por ello, no se visualizan en los ensayos mostrados. \*\*Único evento no compatible con un intrón GT/AG (ver Figura 24). <sup>^</sup>Inserciones que introducen codones de parada en fase, y por lo tanto son eventos PTC-NMD.

Evento <sup>1</sup>	Descripción HGVS <sup>2</sup>	GENCODE <sup>3</sup>	Biotipo <sup>4</sup>	Anotación Funcional <sup>4</sup>	Lecturas <sup>5</sup>	RT-PCR/EC
▼1q <sup>330</sup>	r.145_146ins145+1_145+330	No descrito	Donor shift	No FS	1	E1int-E2int
▼E1A <sup>461</sup> +▼2p <sup>28</sup>	r.145_146ins145+563_145+1023+r.145_146ins146-28_146-1	No descrito	mixed	No FS	1	E1int-E2int
▼3A (alternative 3' end)	r.571_572ins571+3016_571+3104+r.572_1131del	ENST00000622327.4	Terminal modification	intronic STOP+polyA	14	No testado
Δ8,9+▼10	r.966_1131del+r.965+12486_r.965+?	No descrito	mixed	FS-alternative stop	1	No testado

**Tabla 11. Eventos de *splicing* alternativo en *RAD51C* no confirmados.** Los datos de RNAseq sugieren la presencia de al menos 4 eventos de *splicing* alternativo no validados por RT-PCR/EC (dos de ellos no han sido testados por RT-PCR/EC).<sup>1</sup> En Materiales y Métodos se detalla el sistema de designación de eventos. <sup>2</sup> Se siguen las reglas publicadas por la *Human Genome Variation Society* (<http://vrnaomen.hgvs.org/>). <sup>3</sup> Se utilizó Ensembl (<http://www.ensembl.org>) para comprobar si el evento aparece descrito en la compilación 26 de GENCODE. <sup>4</sup> Biotipo y Anotación Funcional según los criterios desarrollados por GENCODE (Mudge et al, 2013). <sup>5</sup> Se muestra el número de lecturas obtenido en el experimento realizado en LCL + puro (media de los valores obtenidos con STAR y TopHat). En el mismo experimento, el número de lecturas (media de STAR y TopHat) correspondiente a cada uno de los 8 intrones de referencia osciló entre 849 y 1654 (media 1249). La cobertura alcanzada en los experimentos realizados en otras muestras biológicas fue sensiblemente inferior, de tal modo que el valor medio de un intrón de referencia fue 291 en la muestra LCL+puro, 358 en Mama, y 52 en fimbria ovárica.

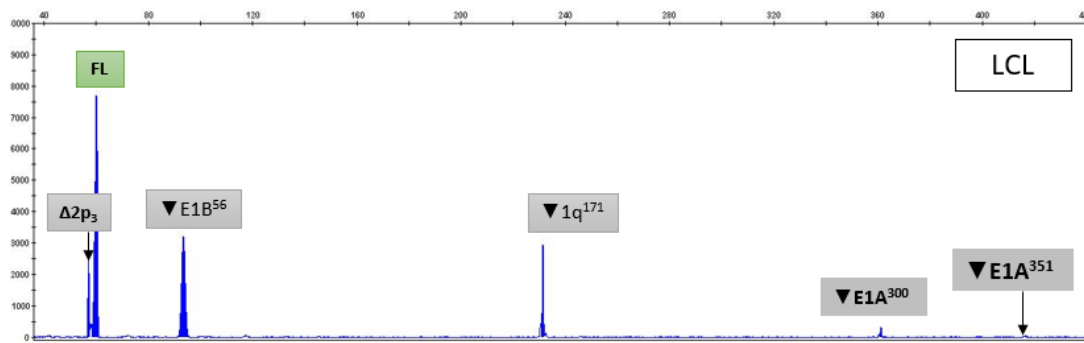
Evento	LCLs	Sangres	Mama	Ovario
$\Delta 1q_{156}$				
$\Delta 1q_{156} + \nabla E1A^{300}$				
$\Delta 1q_{103}$				
$\Delta 1q_{103} + \nabla E1A^{300}$				
$\Delta 1q_{103} \Delta 2p_3$				
$\Delta 1q_{103} \Delta 2$				
$\nabla 1q^{171}$				
$\nabla 1q^{330}$				
$\nabla E1A^{300}$				
$\nabla E1A^{351}$				
$\nabla E1A^{461}$				
$\nabla E1A^{300} + \nabla E1B^{56}$				
$\nabla E1B^{56}$				
$\nabla 2p^{28}$				
$\nabla E1A^{461} + \nabla 2p^{28}$				
$\nabla E1A^{461} + \Delta 2p_3$				
$\Delta 2p_3$				
$\Delta 2$				
$\Delta 2q_{175}$				
$\Delta 2q_{22}$				
$\nabla 2q^{27}$				
$\Delta 3$				
$\Delta 3,4$				
$\Delta 3\_8$				
$\Delta 3q_{114}$				
$\Delta 3q_{114} \Delta 4p_{81}$				
$\nabla 3A$ (alternative 3' end)			n.t.	n.t.
$\Delta 4p_{81}$				
$\Delta 4p_{111}$				

Evento	LCLs	Sangres	Mama	Ovario
$\Delta 4$				
$\Delta 4,5$				
$\Delta 4 + \nabla E4A^{145}$				
$\nabla E4A^{145}$				
$\nabla E4B^{120}$				
$\nabla E4C^{48}$				
$\Delta 5$				
$\Delta 5q_{15/14\_7p_{34/35}}$				
$\nabla E5A^{57}$				
$\nabla E5B^{75}$				
$\nabla E5A^{57} + \nabla E5C^{100}$				
$\nabla E5A^{57} + \nabla E5C^{48}$				
$\nabla E5C^{100}$				
$\nabla E5C^{48}$				
$\Delta 6$				
$\Delta 6,7$				
$\Delta 6,7_{128} + \nabla 8p^3$				
$\Delta 6\_8$				
$\nabla 7p^{22}$				
$\Delta 7$				
$\Delta 7 + \nabla 8p^3$				
$\Delta 7,8$				
$\nabla E7A^{72}$				
$\nabla E7A^{72} + \nabla 8p^3$				
$\Delta 8,9 + \nabla 10$		n.t.	n.t.	n.t.
$\nabla E7B^{122}$				
$\nabla E7B^{122} + \nabla 8p^3$				
$\nabla 8p^3$				
$\Delta 8$				
$\nabla 9p^{31}$				
$\Delta 9p_6$				

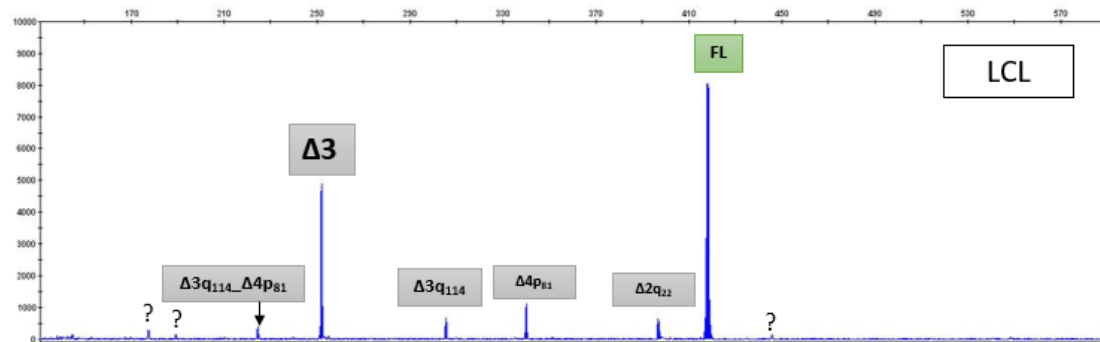
**Tabla 12. Detección de eventos de *splicing* alternativo en *RAD51C* (confirmados y no confirmados) estratificados por tipo de muestra biológica.** Los recuadros sombreados indican que el evento ha sido detectado por RNAseq y/o RT-PCR/EC, mientras que los recuadros blancos indican que el evento no ha sido detectado. La columna Ovario combina los datos obtenidos en Ovario mediante RT-PCR/EC, y los datos obtenidos en fimbria ovárica mediante RNAseq. \*n.t. hace referencia aquellos eventos que no han sido testados por RT-PCR/EC. Muchos eventos minoritarios sólo han sido detectados en LCLs. Probablemente, este hecho simplemente refleje la mayor profundidad del análisis realizado en LCLs. Desde un punto de vista clínico, lo más relevante del análisis es la no detección de eventos de *splicing* alternativo específicos de mama u ovario.

## 2.1. Ensayos RT-PCR/EC realizados en el locus *RAD51C*

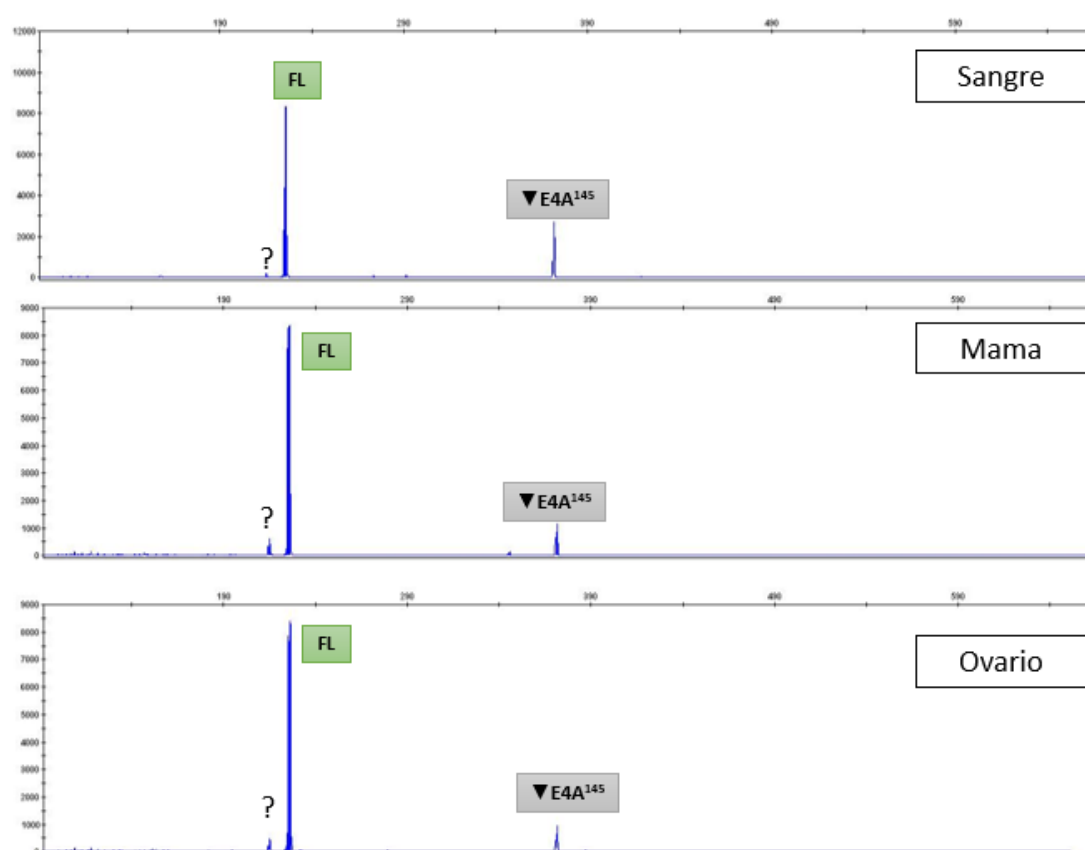
A continuación, se muestran ejemplos representativos de ensayos RT-PCR/EC realizadas a partir muestras de sangre (N=10), LCLs (N=8), tejidos de mama (N=1; 5 réplicas técnicas) y ovario (N=1; 5 réplicas técnicas), y su relación con los datos de RNAseq. En gris aparecen representados los eventos confirmados (detectados por RNAseq y EC), y en naranja los eventos no confirmados (detectados sólo por EC). En verde se representa el transcrito de referencia.



**Figura 21. Ejemplo representativo de ensayo *RAD51C* E1 (int) E2 (int) analizado en una muestra de LCL.** Se detecta el transcrito de referencia (FL, 62nt),  $\Delta 2p_3$  (59nt),  $\nabla E1B^{56}$  (118nt),  $\nabla 1q^{171}$  (233nt),  $\nabla E1A^{300}$  (362nt) y  $\nabla E1A^{351}$  (413nt). Todos los eventos de *splicing* alternativo se detectan con una intensidad de señal menor que la del transcrito de referencia.

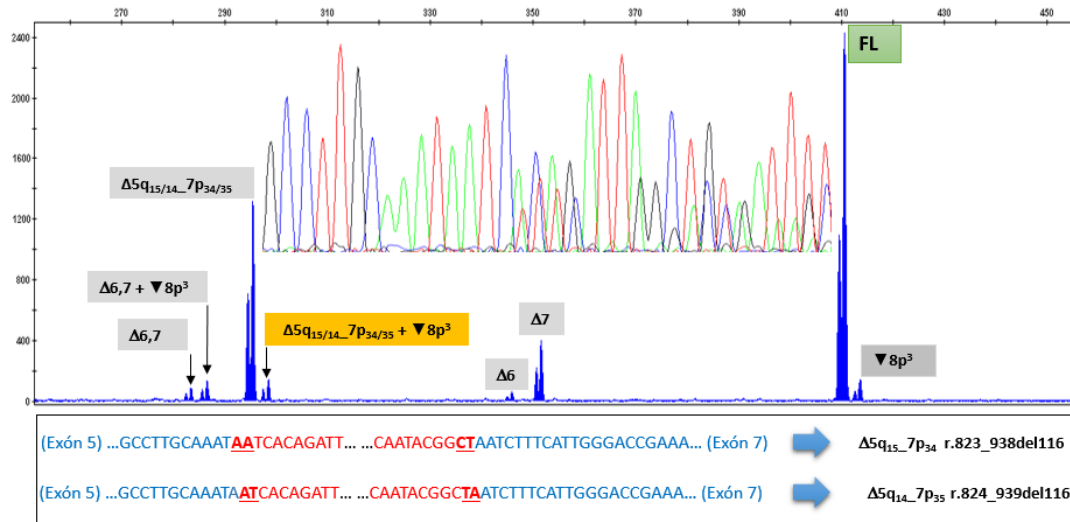


**Figura 22. Ejemplo representativo de ensayo *RAD51C* E2E4 analizado en una LCL.** Se detecta el transcrito de referencia (FL, 419nt),  $\Delta 2q_{22}$  (397nt),  $\Delta 4p_{81}$  (338nt),  $\Delta 3q_{114}$  (305nt),  $\Delta 3$  (252nt) y  $\Delta 3q_{114}\text{--}\Delta 4p_{81}$  (224nt). Con un signo de interrogación se representan los picos para los que no hemos podido inferir ninguna anotación. Todos los eventos de *splicing* alternativo se detectan como productos de menor intensidad que el transcrito de referencia. Con el signo de interrogación (?) señalamos potenciales eventos de *splicing* no anotados (no hemos sabido inferir ninguna anotación en base a la diferencia de tamaño con el transcrito de referencia).

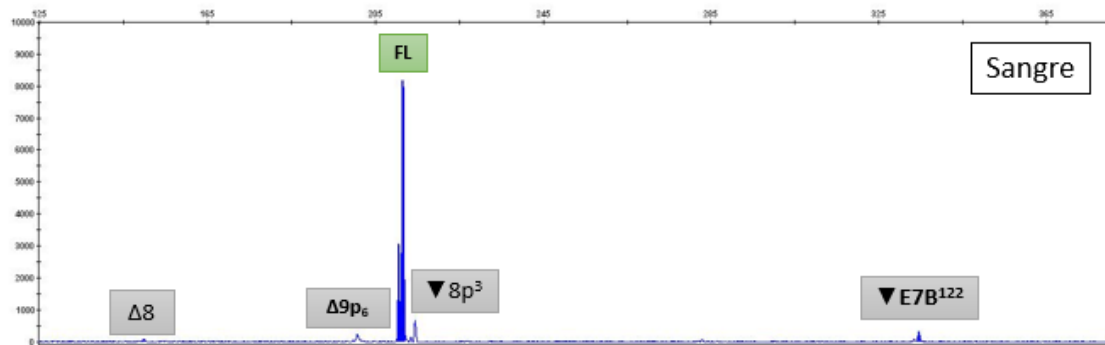


**Figura 23. Ejemplos representativos de ensayo *RAD51C* E4E5 en sangre, mama y ovario.** Se detectan el transcrito de referencia (FL, 226nt) y  $\nabla$ E4A<sup>145</sup> (371nt). Con el signo de interrogación (?) señalamos un potencial evento de *splicing* no anotado (no hemos sabido inferir ninguna anotación en base a la diferencia de tamaño con el transcrito de referencia).





**Figura 24. Caracterización de un evento de *splicing* alternativo no canónico.** El panel superior muestra un ejemplo representativo de un ensayo *RAD51C* E4E8 realizado en mama. Se detectan el transcrito de referencia (FL, 411nt) y los eventos de *splicing* alternativo  $\Delta 6,7$  (283nt),  $\Delta 6,7 + \nabla 8p^3$  (286nt),  $\Delta 6$  (344nt),  $\Delta 7$  (350nt) y  $\nabla 8p^3$  (414nt). Además, se detecta un producto de 293 nucleótidos muy predominante que anotamos provisionalmente como  $\Delta 5q_{15/14\_7p_{34/35}}$ . Tras realizar una secuenciación Sanger se demuestra que el producto corresponde a un transcrito que carece de los últimos nucleótidos del exón 5, todo el exón 6, y los primeros nucleótidos del exón 7. El evento no puede ser explicado por el reconocimiento de sitios canónicos de *splicing* GT/AG, aunque sí por el reconocimiento de sitios no canónicos AA/CT o AT/TA (Parada et al, 2014; Sibley et al, 2016). Dependiendo de si estamos ante un intrón AA/CT o AT/TA, la anotación varía, tal y como se esquematiza en el panel inferior (intrón atípico resaltado en rojo). No tenemos lecturas de RNAseq compatibles con  $\Delta 5q_{15/14\_7p_{34/35}}$ , pero lo consideramos un evento confirmado en base a la secuenciación Sanger.

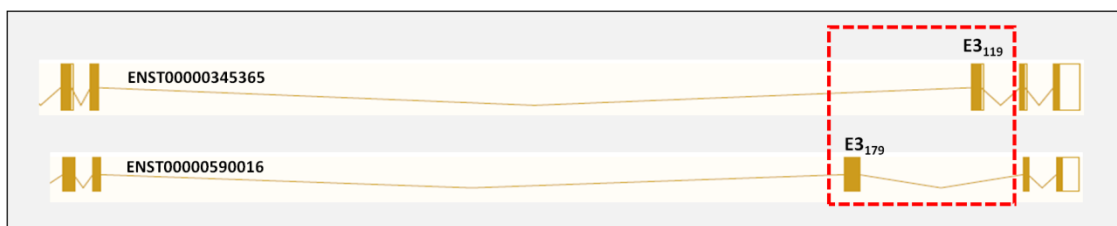


**Figura 25. Ejemplo representativo de ensayo *RAD51C* E7E9 realizado en una muestra de sangre.** Se detecta el transcrito de referencia (FL, 211nt); y los eventos  $\Delta 8$  (150nt),  $\Delta 9p_6$  (205nt),  $\nabla 8p^3$  (214nt), y  $\nabla E7B^{122}$  (333nt). Todos los eventos de *splicing* alternativo se detectan con una intensidad de señal menor que la del transcrito de referencia.

### 3. Caracterización del *splicing* alternativo en el gen de susceptibilidad *RAD51D*

#### 3.1. Identificación del exón 3 de referencia

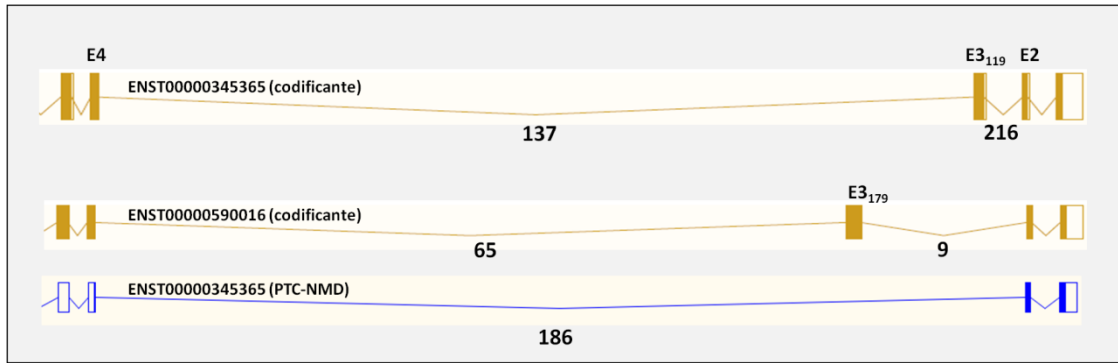
Tal y como se muestra en la figura 26, la selección de un transcrito de referencia en el gen *RAD51D* no es una tarea obvia. De hecho, en la literatura científica no existe consenso en cuanto al exón 3 de referencia (Loveday et al., 2011; Lu C et al., 2015), lo que provoca inconsistencia en la anotación de variantes genéticas.



**Figura 26. El locus *RAD51D* contiene dos exones 3 mutuamente excluyentes.** Tal y como se muestra en la figura, los transcritos ENST00000345365 y ENST00000590016 son idénticos, salvo por la presencia de dos exones 3 alternativos y mutuamente excluyentes de 119 y 179 nucleótidos. La selección de un transcrito de referencia no es obvia. Figura adaptada de Ensembl.

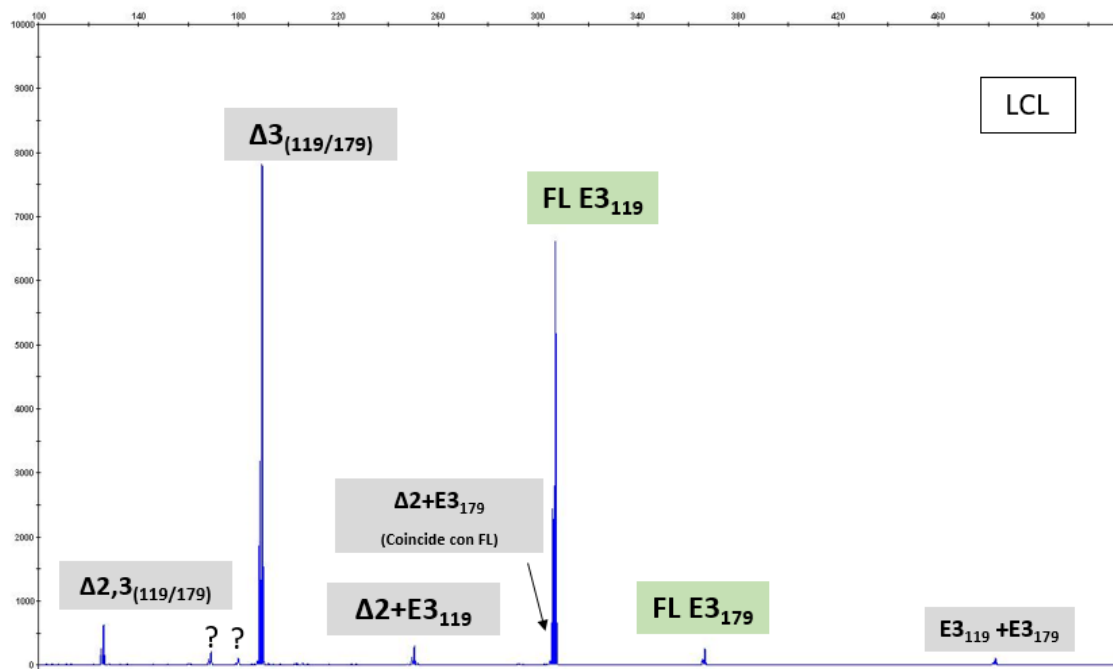
Por ello, iniciamos la caracterización del *splicing* alternativo en *RAD51D* determinando si los niveles de expresión de los dos exones 3 alternativos eran muy distintos, justificando así la selección del mayoritario como transcrito de referencia.

En primer lugar, atendiendo al número de lecturas de RNAseq (Figura 27), podemos ver que se registran mayor número de lecturas para el transcrito **ENST00000345365** que contiene el exón 3 de 119 nucleótidos. Es interesante señalar que el número de lecturas correspondientes a transcritos sin exón 3 también es muy elevado (más elevado de hecho que el número de lecturas que justifican el exón 3 de 179 nucleótidos).



**Figura 27. Análisis por RNAseq del exón 3 de *RAD51D*.** Se muestran 3 transcritos correspondientes a la inclusión del exón 3 de 119 nucleótidos, el exón 3 de 179 nucleótidos, y la no inclusión de exón 3 (PTC-NMD). La representación gráfica de los 3 transcritos y los códigos identificativos se han obtenido de Ensembl. La figura muestra las lecturas de RNAseq correspondientes a los intrones E2E3<sub>119</sub> (216), E3<sub>119</sub>E4 (137), E2E3<sub>179</sub> (9), E3<sub>179</sub>E4 (65), y E2E4 (186). El número de lecturas corresponde a la media de lecturas obtenidas en LCL + puro con los algoritmos STAR y TopHat.

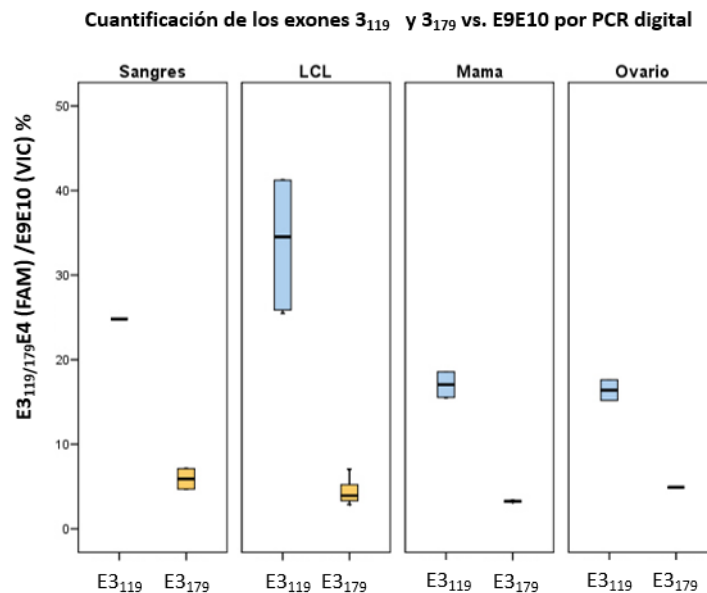
En segundo lugar, los datos obtenidos por RT-PCR coinciden en: (1) es más predominante el exón 3 de 119 nucleótidos, y (2) en la alta proporción de transcritos sin exón 3.



**Figura 28. Ejemplo representativo de ensayo *RAD51D* E1E4 realizado en una LCL.** Se puede observar cómo el transcrito de referencia del exón 3 de 119 nucleótidos (FL E3<sub>119</sub>) se expresa con más intensidad de señal que el transcrito de referencia del exón 3 de 179 nucleótidos (FL E3<sub>179</sub>). También es muy destacable la intensidad con la que se detectan transcritos sin ninguno de los dos exones 3. Además, se detectan transcritos que demuestran que los dos exones 3 de *RAD51D* no son mutuamente excluyentes (pico de 483nt). La coexistencia de los dos exones 3 en un único transcrito se confirmó en el ensayo RT-PCR/EC E3<sub>119</sub>-E4 (datos no mostrados). Los datos de RNAseq también respaldan este punto.

En tercer lugar, el carácter mayoritario del exón 3<sub>119</sub> se ha confirmado mediante PCR digital (dPCR). Para ello se han diseñado dos ensayos TaqMan marcados con FAM. Uno que reconoce la unión del exón 3<sub>119</sub> con el exón 4. Y otro que reconoce la unión exón E3<sub>179</sub> con el exón 4. Como referencia se ha empleado un ensayo TaqMan marcado con VIC que reconoce la unión de los exones 9 y 10 de *RAD51D*.

Los resultados obtenidos por PCR digital indican que es mayor la contribución de E3<sub>119</sub> en todos los tipos de muestra, con respecto a la de E3<sub>179</sub>. El E3<sub>179</sub> parece que se expresa de forma equivalente en todas las formas de tejido, en torno al 5%. Mientras que la contribución del exón E3<sub>119</sub> es del 20% en los tejidos de mama y ovario, cerca del 25% en Sangres, y en LCLs entorno al 35%. Es decir, de forma global es más predominante E3<sub>119</sub>, por lo que se ha considerado como transcrito de referencia aquel que contiene el exón 3 de 119 nucleótidos.



**Figura 29. Diagrama de cajas que muestra la contribución relativa de los exones E3<sub>179</sub> y E3<sub>119</sub> a la expresión global de *RAD51D* (E9E10) por dPCR.** Se muestran los datos agrupados de cada una de las muestras y según el exón cuantificado. Para la cuantificación del exón 3<sub>119</sub>: Sangres (n=1), LCLs (n=6), Tejido de Mama (n=3) y tejido de Ovario (n=2). Y en la cuantificación del exón 3<sub>179</sub> se han empleado: Sangres (n=1), LCLs (n=6), Tejido de Mama (n=1) y tejido de Ovario (n=1). En el caso de los tejidos de mama y ovario la N se refiere a réplicas técnicas. Los diagramas muestran el valor mínimo, percentil 25%, mediana, percentil 75% y valor máximo.

En resumen, los datos de RNAseq, RT-PCR/EC y PCR digital coinciden en señalar que los niveles de expresión del exón E3<sub>119</sub> son superiores a los del E3<sub>179</sub>, al menos en las muestras biológicas analizadas. Por lo que, hemos decidido describir el *splicing* alternativo en *RAD51D* considerando a **ENST00000345365** como transcrito de referencia.

### 3.2. Anotación de eventos de *splicing* alternativo

En total se han confirmado 30 eventos de *splicing* alternativo en *RAD51D* (Tabla 13), de los cuales 20 son truncantes (PTC-NMD), 9 mantienen la fase de lectura (No Fs), y 1 genera un codón de stop alternativo (Fs-Alternative stop). Otros eventos no confirmados se describen en la Tabla 14.

Evento <sup>1</sup>	Descripción HGVS <sup>2</sup>	GENCODE <sup>3</sup>	Biotipo <sup>4</sup>	Anotación funcional <sup>4</sup>	Lecturas <sup>5</sup>	RT-PCR/EC
$\Delta 2,3$	r.83_263del	No descrito	multicasette	No Fs	5,5	Figura 28
$\Delta 2_4$	r.83_345del	No descrito	multicasette	PTC-NMD	1	PCR E1E9
$\Delta 3,4 + \nabla 5p^{182}$	r.145_345del+r.345_346ins346-182_346-1	No descrito	multicasette+acceptor shift	PTC-NMD	1	PCR E1E5
$\Delta 2_5$	r.83_480del	No descrito	multicasette	PTC-NMD	1	PCR E1E9
$\Delta 3$	r.145_263del	ENST00000345365	casette	PTC-NMD	186	Figura 28, 31, 32
$\Delta 3,4$	r.145_345del	ENST00000345365	multicasette	No Fs	5	PCR E1E5/PCR E2E8*
$\Delta 3_5$	r.145_480del	ENSG00000185379	multicasette	No Fs	61	Figura 28-32
$\Delta 3_6p_4$	r.145_484del	No descrito	multicasette+acceptor shift	PTC-NMD	2,5	PCR E2E8*
$\Delta 3_6$	r.145_576del	ENST00000345365	multicasette	No Fs	4,5	PCR E2E8, E1E9, E1E7
$\Delta 3 + \nabla E3A^{179}$	r.145_263del+r.263_264ins263+1464_263+1642	ENST00000590016	casette+casette	No Fs	37	Figura 28
$\Delta 3 + \nabla E3A^{179} + \Delta 4,5$	r.145_263del+r.263_264ins263+1464_263+1642+r.264_480del	ENST00000592850.5	casette+casette+multicasette	PTC-NMD	28	PCR E2E8*
$\Delta 3 + \nabla E3B^{98}$	r.145_263del+r.263_264ins263+1709_263+1806	No descrito	casette + casette	No Fs	1	PCR E1E5*
$\nabla E3A^{179}$	r.263_264ins263+1464_263+1642	ENST00000587977	casette	PTC-NMD	1	PCR E3 <sub>119</sub> _E4
$\nabla E3A^{179} + \Delta 4,5$	r.263_264ins263+1464_263+1642+r.264_480del	No descrito	mixed	PTC-NMD	7,5	PCR E2E8*
$\nabla E3A^{179} + \Delta 4_6p_4$	r.263_264ins263+1464_263+1642+r.264_484del	No descrito	mixed	PTC-NMD	1	PCR E3 <sub>179</sub> _E7
$\Delta 4$	r.264_345del	ENST00000586210	casette	PTC-NMD	5,5	Figura 33
$\Delta 4,5$	r.264_480del	ENST00000592928.2	multicasette	PTC-NMD	26	Figura 30, 32
$\Delta 4_6p_4$	r.264_484del	No descrito	multicasette+acceptor shift	PTC-NMD	4	PCR E2E8*

Evento <sup>1</sup>	Descripción HGVS <sup>2</sup>	GENCODE <sup>3</sup>	Biotipo <sup>4</sup>	Anotación funcional <sup>4</sup>	Lecturas <sup>5</sup>	RT-PCR/EC
Δ4q <sub>17</sub>	r.329_345del	No descrito	donor shift	PTC-NMD	14 <sup>#</sup>	Figura 33
Δ5p <sub>61</sub>	r.346_406del	ENST00000460118	acceptor shift	PTC-NMD	34 <sup>#</sup>	PCR E2E8*
Δ5p <sub>99</sub>	r.346_444del	No descrito	acceptor shift	No Fs	4	PCR E2E8*
Δ5	r.346_480del	ENST00000588594	cassette	No Fs	30	Figura 34
Δ5,6p <sub>4</sub>	r.346_484del	No descrito	cassette+acceptor shift	PTC-NMD	3	PCR E2E8*
Δ6p <sub>4</sub>	r.481_484del	No descrito	acceptor shift	PTC-NMD	1	PCR E2E8*
▼E6A <sup>163</sup>	r.576_577ins576+96_576+258	No descrito	cassette	PTC-NMD	1	PCR E6E7*
▼E6A <sup>122</sup>	r.576_577ins576+137_576+258	ENST00000588372	cassette	PTC-NMD	58	Figura 35
▼E6A <sup>55</sup>	r.576_577ins576+204_576+258	No descrito	cassette	PTC-NMD	1	Figura 35
Δ7,8	r.577_738del	No descrito	multicassette	No Fs	3	Figura 36
▼9p <sup>16</sup>	r.738_739ins739-16	No descrito	acceptor shift	PTC-NMD	1	PCR E6E9*
Δ10p <sub>7</sub>	r.904_910del	No descrito	acceptor shift	FS-alternative stop	2	PCR E6E10

**Tabla 13. Eventos de *splicing* alternativo en *RAD51D* confirmados.** <sup>1</sup> En Materiales y Métodos se detalla el sistema de designación de eventos. <sup>2</sup> Se siguen las reglas publicadas por la *Human Genome Variation Society* (<http://vrnaomen.hgvs.org/>). <sup>3</sup> Se utilizó Ensembl (<http://www.ensembl.org>) para comprobar si el evento aparece descrito en la compilación 26 de GENCODE. <sup>4</sup> Biotipo y Anotación Funcional según los criterios desarrollados por GENCODE (Mudge et al, 2013). <sup>5</sup> Se muestra el número de lecturas obtenido en el experimento realizado en LCL+puro (media de los valores obtenidos con STAR y TopHat). En el mismo experimento, el número de lecturas (media de STAR y TopHat) correspondiente a cada uno de los 9 intrones de referencia osciló entre 137 y 551 (media 400). La cobertura alcanzada en los experimentos realizados en otras muestras biológicas fue sensiblemente inferior, de tal modo que el valor medio de un intrón de referencia fue 26 en la muestra LCL-puro, 218 en Mama, y 20 en fimbria ovárica. \*Eventos que han sido detectados en los ensayos señalados, pero se detectan como productos de muy baja intensidad. Y sólo se detectan en algunas muestras. Por ello no se visualizan en los ensayos mostrados. <sup>#</sup>Se destacan eventos con número de lecturas muy discrepantes en STAR y TopHat (por ejemplo, Δ4q<sub>17</sub> tiene 3 lecturas en STAR y 25 en TopHat).

Eventos <sup>1</sup>	HGVS <sup>2</sup>	Gencode <sup>3</sup>	Biotipo <sup>4</sup>	Anotación Funcional <sup>4</sup>	Lecturas <sup>5</sup>	RT-PCR/EC
$\Delta 2p_{21}$ (sólo detectado en transcritos $\Delta 3$ )	r.83_103del	No descrito	acceptor shift	No Fs	-	RT-PCR/EC E1E4*
$\Delta 2$	r.83_144del	No descrito	cassette	PTC-NMD	-	RT-PCR/EC E1E4*
$\Delta 2q_{10\_3}$	r.135_263del	No descrito	donor shift+cassette	PTC-NMD	-	RT-PCR/EC E1E5
$\Delta 3 + \nabla E3A^{179} + \Delta 4\_6$	r.145_576del+r.263_264ins263+1464_263+1642	No descrito	multicassette+cassette	No Fs	-	RT-PCR/EC E2E8*
$\Delta 6$	r.481_576	No descrito	cassette	No FS	-	RT-PCR/EC E2E8*
$\Delta 7$	r.577_667del	No descrito	cassette	No FS	-	Figura 32
$\Delta 7\_9p_{108}$	r.577_846del	No descrito	multicassette	No Fs	-	RT-PCR/EC E6E10
$\nabla E3B^{98}$	r.263_264ins263+1709_263+1806	No descrito	cassette	PTC-NMD	47,25	No testado
$\nabla E3B^{98} + \Delta 4$	r.263_264ins263+1709_263+1806+r.264_345del	No descrito	cassette+cassette	PTC-NMD	1	E2E8
$\Delta 4\_6$	r.246_576del	No descrito	multicassette	PTC-NMD	5,5 <sup>#</sup>	E2E8
$\Delta 10p_{145}$	r.904_*50del	No descrito	acceptor shift	FS-alternative stop	1	E6E10

**Tabla 14. Eventos de *splicing* alternativo en *RAD51D* no confirmados.** Los datos de RNAseq (señalados en amarillo) sugieren la presencia de al menos 4 eventos de *splicing* alternativo no validados por RT-PCR/EC (uno de ellos no han sido testados por RT-PCR/EC).<sup>1</sup> Los datos de RT-PCR/EC (señalados en naranja) sugieren la presencia de al menos 7 eventos de *splicing* alternativo no validados por RNAseq. En Materiales y Métodos se detalla el sistema de designación de eventos.<sup>2</sup> Se siguen las reglas publicadas por la *Human Genome Variation Society* (<http://vrnaomen.hgvs.org/>).<sup>3</sup> Se utilizó Ensembl (<http://www.ensembl.org>) para comprobar si el evento aparece descrito en la compilación 26 de GENCODE.<sup>4</sup> Biotipo y Anotación Funcional según los criterios desarrollados por GENCODE (Mudge et al, 2013).<sup>5</sup> Se muestra el número de lecturas obtenido en el experimento realizado en LCL+puro (media de los valores obtenidos con STAR y TopHat). En el mismo experimento, el número de lecturas (media de STAR y TopHat) correspondiente a cada uno de los 9 intrones de referencia osciló entre 137 y 551 (media 400). La cobertura alcanzada en los experimentos realizados en otras muestras biológicas fue sensiblemente inferior, de tal modo que el valor medio de un intrón de referencia fue 26 en la muestra LCL-puro, 218 en Mama, y 20 en fimbria ovárica. \*Eventos que han sido detectados en los ensayos señalados, pero se detectan como productos de muy baja intensidad. Y sólo se detectan en algunas muestras. Por ello no se visualizan en los ensayos mostrados. <sup>#</sup>Se destacan eventos con número de lecturas muy discrepantes en STAR y TopHat (por ejemplo,  $\Delta 4\_6$  tiene 1 lecturas en STAR y 10 en TopHat).

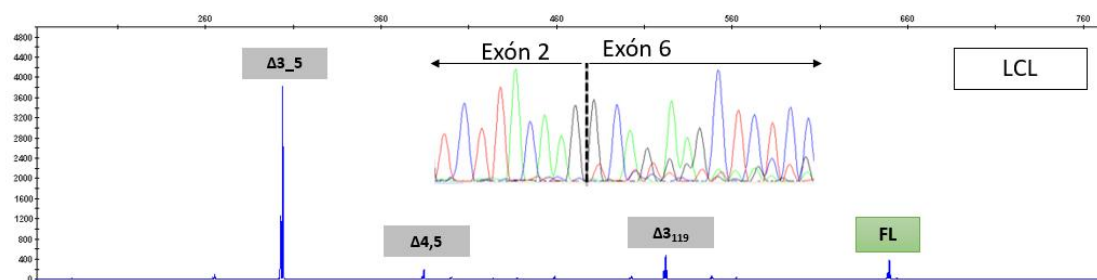
Evento	LCLs	Sangres	Mama	Ovario	Evento	LCLs	Sangres	Mama	Ovario
$\Delta 5'UTR_{1p140}$	n.t.	n.t.		n.t.	$\nabla E3A^{179} + \Delta 4_{6p4}$				
$\Delta 5'UTR_1$		n.t.	n.t.	n.t.	$\Delta 4$				
$\Delta 5'UTR_3$		n.t.	n.t.	n.t.	$\Delta 4,5$				
$\Delta 1q_{168}$		n.t.	n.t.	n.t.	$\Delta 4_{6p4}$				
$\Delta 2p_{21}$					$\Delta 4_6$				
$\Delta 2$					$\Delta 4q_{17}$				
$\Delta 2,3$					$\Delta 5p_{61}$				
$\Delta 2_4$					$\Delta 5p_{99}$				
$\Delta 3,4 + \nabla 5p^{182}$					$\Delta 5$				
$\Delta 2_5$					$\Delta 5,6p_4$				
$\Delta 2q_{10\_3}$					$\Delta 6p_4$				
$\Delta 3$					$\Delta 6$				
$\Delta 3,4$					$\nabla E6A^{163}$				
$\Delta 3_5$					$\nabla E6A^{122}$				
$\Delta 3_{6p4}$					$\nabla E6A^{55}$				
$\Delta 3_6$					$\Delta 7$				
$\Delta 3 + \nabla E3A^{179}$					$\Delta 7,8$				
$\Delta 3 + \nabla E3A^{179} + \Delta 4,5$					$\Delta 7_9p_{108}$				
$\Delta 3 + \nabla E3A^{179} + \Delta 4_6$					$\nabla 9p^{16}$				
$\Delta 3 + \nabla E3B^{98}$					$\Delta 10p_7$				
$\nabla E3A^{179}$					$\Delta 10p_{145}$				
$\nabla E3B^{98}$					$\Delta 10int_{1086}$		n.t.	n.t.	n.t.
$\nabla E3B^{98} + \Delta 4$					$\Delta 10int_{916}$		n.t.	n.t.	n.t.
$\nabla E3A^{179} + \Delta 4,5$									

**Tabla 15. Detección de eventos de *splicing* alternativo en *RAD51D* (confirmados y no confirmados) estratificados por tipo de muestra biológica.** Los recuadros sombreados indican que el evento ha sido detectado por RNAseq y/o RT-PCR/EC, mientras que los recuadros blancos indican que el evento no ha sido detectado. La columna Ovario combina los datos obtenidos en Ovario mediante RT-PCR/EC, y los datos obtenidos en fimbria ovárica mediante RNAseq. \*n.t. hace referencia aquellos eventos que no han sido testados por RT-PCR/EC. Muchos eventos minoritarios sólo han sido detectados en LCLs. Probablemente, este hecho simplemente refleje la mayor profundidad del análisis realizado en LCLs. Desde un punto de vista clínico, lo más relevante del análisis es la no detección de eventos de *splicing* alternativo específicos de mama u ovario.

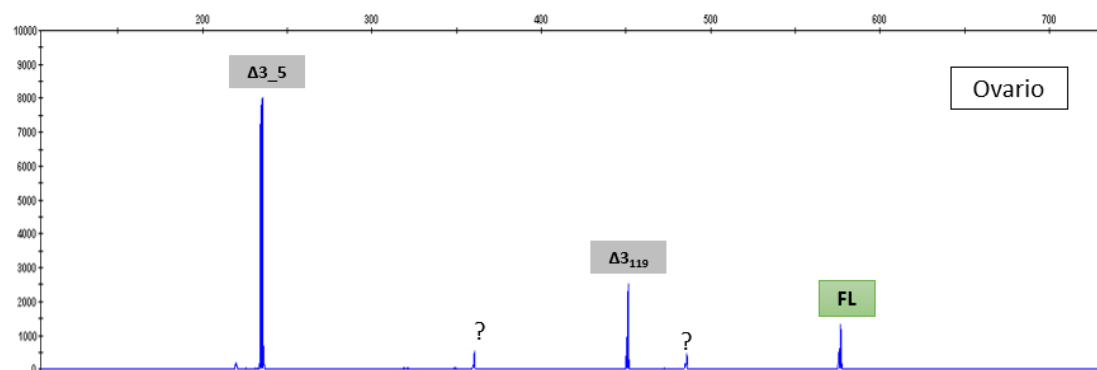


### 3.3. Ensayos RT-PCR/EC realizados en el locus *RAD51D*

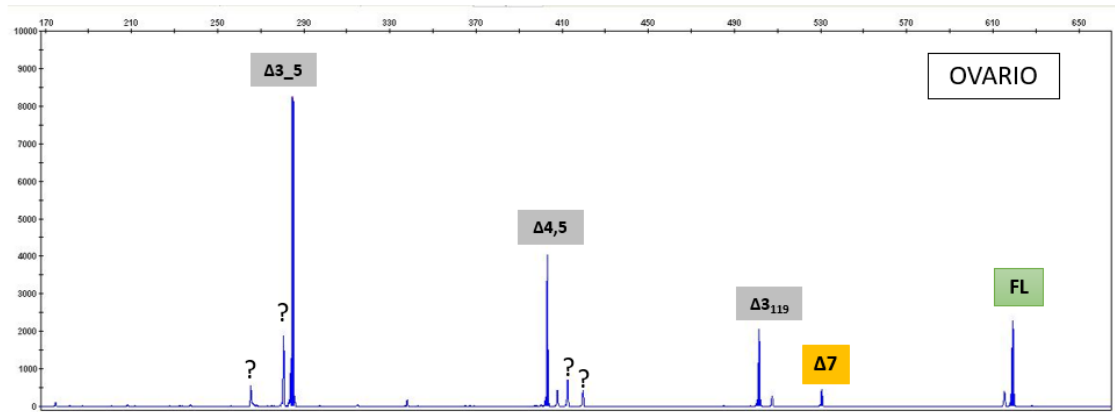
A continuación, se muestran algunos ejemplos representativos de ensayos RT-PCR/EC realizadas a partir muestras de sangre (N=6), LCLs (N=10), tejidos de mama (N=1; 5 réplicas técnicas) y ovario (N=1; 5 réplicas técnicas), y su relación con los datos de RNAseq. En gris aparecen representados los eventos confirmados (detectados por RNAseq y EC), y en naranja los eventos no confirmados (detectados sólo por EC). En verde se representa el transcrito de referencia.



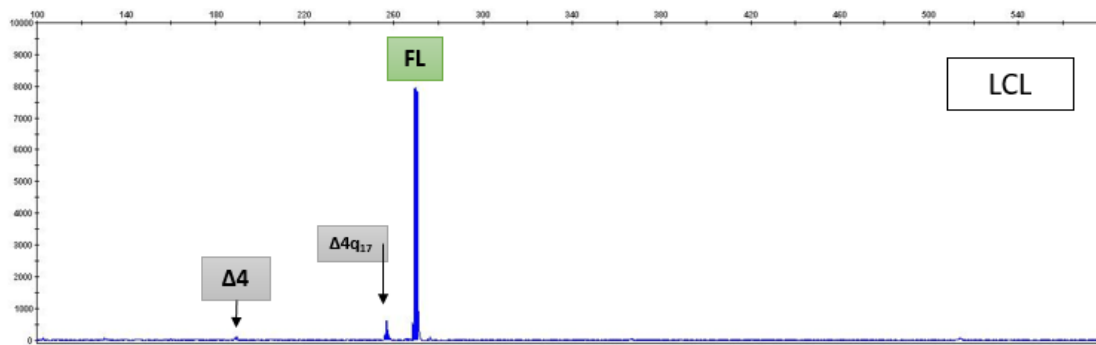
**Figura 30. Ejemplo representativo de ensayo *RAD51D* E1E7 realizado en una muestra LCL.** Se detecta el transcrito de referencia (FL, 650 nt) y las isoformas por *splicing* alternativo  $\Delta 3_5$  (303nt),  $\Delta 4,5$  (384nt) y  $\Delta 3$  (523nt). Este ensayo ha sido analizado empleando como marcador LIZ600. La intensidad relativa del pico correspondiente a  $\Delta 3_5$  permite su caracterización por secuenciación Sanger (electroferograma mostrado). El evento de *splicing*  $\Delta 3_5$  se detecta con una intensidad de señal relativa muy elevada (sobrestimación debida a la amplificación preferencial de fragmentos pequeños). Tras realizar una secuenciación Sanger (Secuenciado con cebador *sense*) se demuestra que el producto corresponde a un transcrito que carece de los exones 3, 4 y 5 ( $\Delta 3_5$ ).



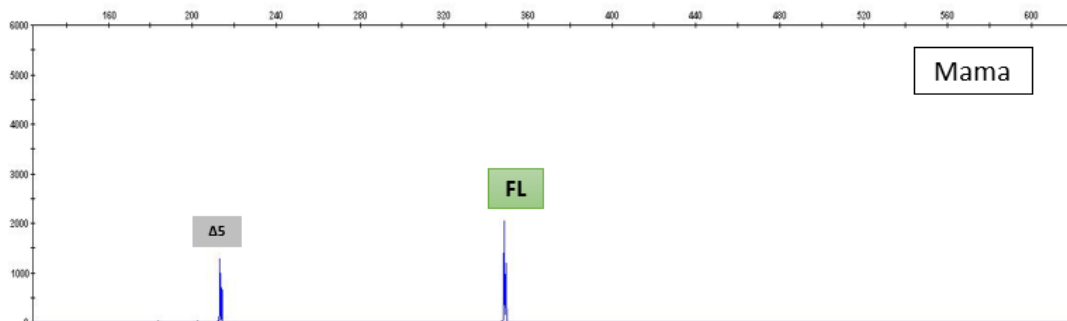
**Figura 31. Ejemplo representativo de ensayo E2E7 *RAD51D* realizado en ovario.** Se detecta el transcrito de referencia (FL, 576nt), y los eventos de *splicing* alternativo  $\Delta 3_5$  (235nt), y  $\Delta 3_{119}$  (451nt). Con el signo de interrogación (?) señalamos potenciales eventos de *splicing* no anotados (no hemos sabido inferir ninguna anotación en base a la diferencia de tamaño con el transcrito de referencia). El evento de *splicing*  $\Delta 3_5$  se detecta con una intensidad de señal relativa muy elevada (sobrestimación debida a la amplificación preferencial de fragmentos pequeños).



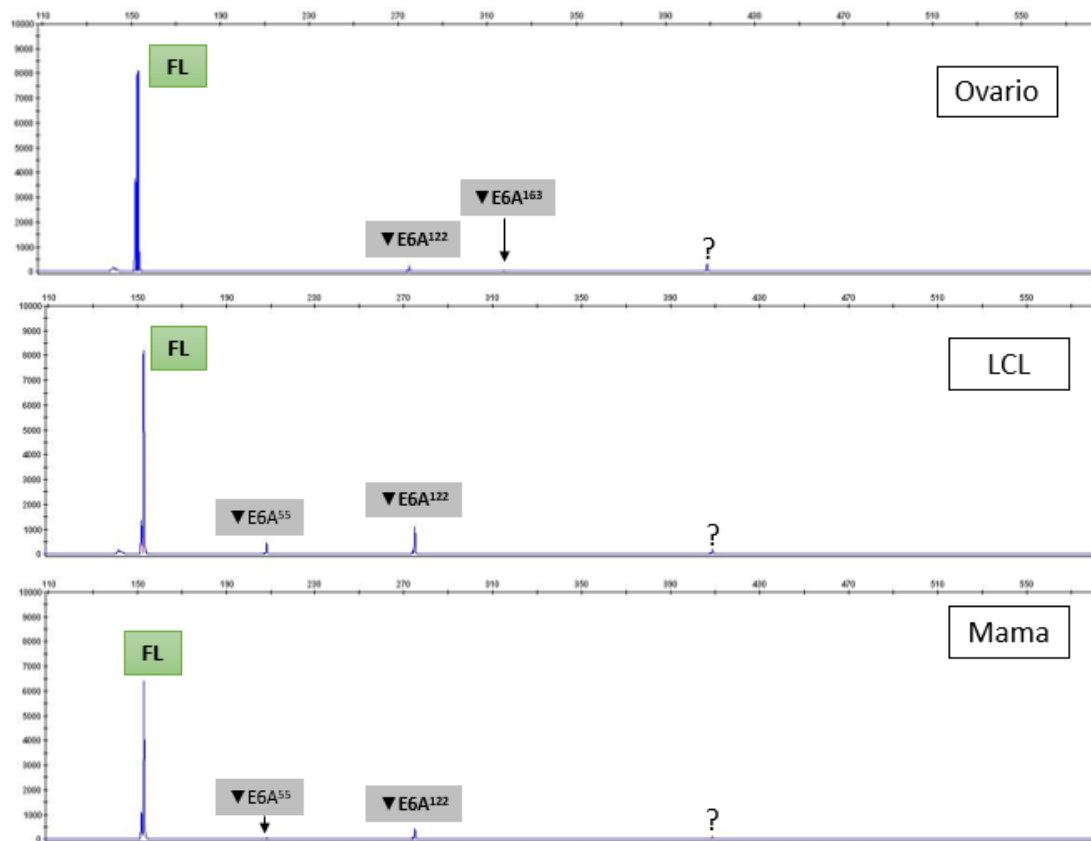
**Figura 32. Ejemplo representativo de ensayo *RAD51D* E2E8 realizado en ovario.** Se detectan el transcrito de referencia (FL, 619nt),  $\Delta 3\_5$  (285nt),  $\Delta 3\_6p_4$  (289nt) y  $\Delta 4,5$  (403nt),  $\Delta 3_{119}$  (501nt),  $\Delta 7$  (531nt) y  $\Delta 6p_4$  (615nt). Con el signo de interrogación (?) señalamos potenciales eventos de *splicing* no anotados (no hemos sabido inferir ninguna anotación en base a la diferencia de tamaño con el transcrito de referencia). Las muestras se analizaron con LIZ600. El evento de *splicing*  $\Delta 3\_5$  se detecta con una intensidad de señal relativa muy elevada (sobrestimación debida a la amplificación preferencial de fragmentos pequeños).



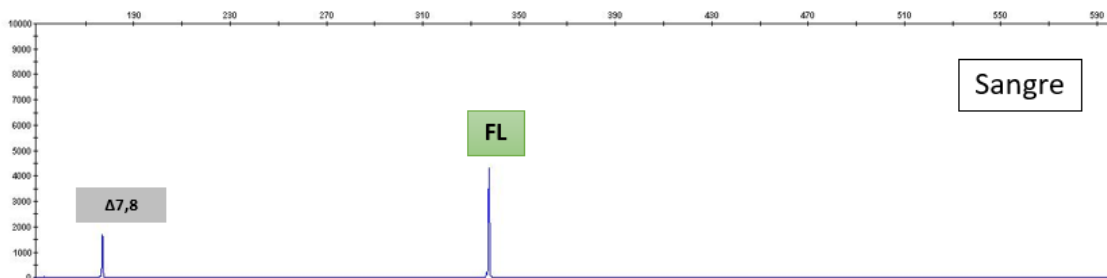
**Figura 33. Ejemplo representativo de ensayo *RAD51D* E3E5 realizado en una LCL.** Se detecta el transcrito de referencia (FL, 271nt), y los eventos de *splicing* alternativo  $\Delta 4$  (189nt) y  $\Delta 4q_{17}$  (257nt). Todos los eventos de *splicing* alternativo se detectan con una intensidad de señal menor que la del transcrito de referencia.



**Figura 34. Ejemplo representativo de ensayo *RAD51D* E4E7 realizado en mama.** Se detecta el transcrito de referencia (FL, 348nt), y  $\Delta 5$  (214nt).



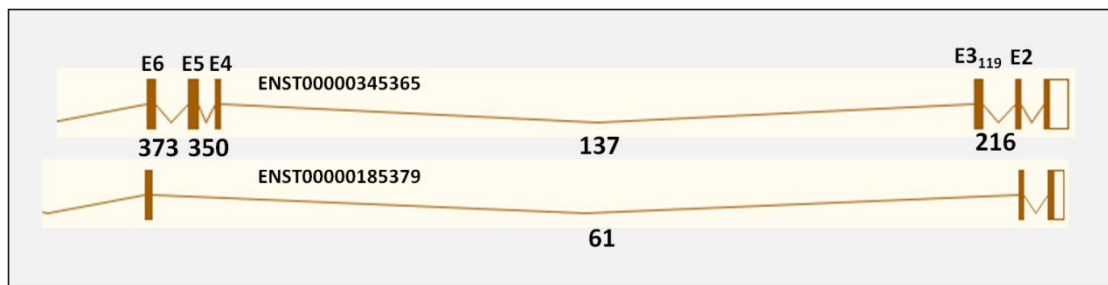
**Figura 35. Ejemplo representativo de ensayo *RAD51D* E6E7 realizados en ovario, una LCL y mama.** Se detectan el transcrito de referencia (FL, 152nt) y los eventos  $\nabla E6A^{55}$  (208nt),  $\nabla E6A^{122}$  (274nt) y  $\nabla E6A^{163}$  (319nt). Con el signo de interrogación (?) señalamos potenciales eventos de *splicing* no anotados (no hemos sabido inferir ninguna anotación en base a la diferencia de tamaño con el transcrito de referencia).



**Figura 36. Ejemplo representativo de ensayo *RAD51D* E6E9 realizado en una muestra de sangre.** Se detecta el transcrito de referencia (337nt) y el evento  $\Delta 7,8$  (177nt).

### 3.4. Contribución del evento de *splicing* $\Delta 3\_5$ sobre la expresión global de *RAD51D*

Los datos de RT-PCR/EC sugieren que  $\Delta 3\_5$  es un evento predominante (aunque probablemente sobreestimado como resultado de la amplificación preferencial de fragmentos pequeños) en los cuatro tipos de muestras analizadas. De mismo modo, los datos de RNAseq sugieren que  $\Delta 3\_5$  es un evento de *splicing* alternativo predominante (Figura 37).

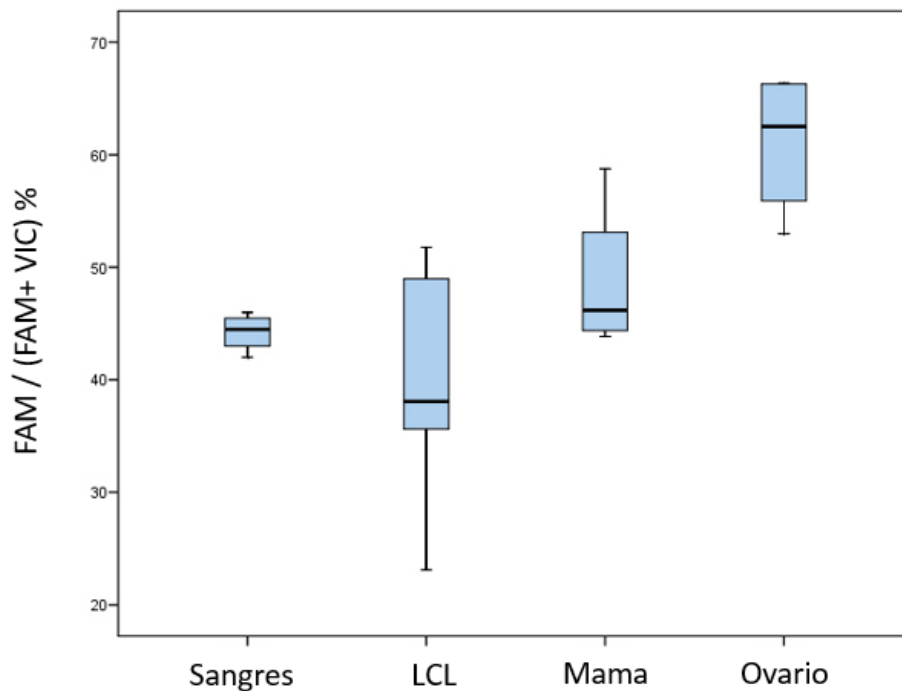


**Figura 37. Análisis por RNAseq de *RAD51D*  $\Delta 3\_5$ .** Se muestra el transcrito de referencia con el exón 3 de 119 nucleótidos, y el transcrito  $\Delta 2\_5$ . La representación gráfica de los 2 transcritos y los códigos identificativos se han obtenido de Ensembl. La figura muestra las lecturas de RNAseq correspondientes a los intrones E2E3<sub>119</sub> (216), E3<sub>119</sub>E4 (137), E4E5 (350), E5E6 (373), y E2E6 (61). El número de lecturas corresponde a la media de lecturas obtenidas en LCL + puro con los algoritmos STAR y TopHat.

Dada la importancia potencial de  $\Delta 3\_5$  para la clasificación clínica de variantes (es un evento que mantiene la fase de lectura), decidimos cuantificar por PCR digital la contribución de este evento de *splicing* alternativo sobre la expresión global de *RAD51D*. Para ello, se han diseñado dos ensayos TaqMan. Por un lado, un ensayo TaqMan marcado con FAM, que reconoce la unión de los exones 2 y 6 de *RAD51D*. Y por otro lado, un ensayo TaqMan marcado con VIC que reconoce la unión de los exones 3 y 4 de *RAD51D* (ver métodos).

La expresión de mensajeros que contienen  $\Delta 3\_5$  se acerca al 40% en muestras de sangre, LCLs, y mama, y supera el 70% en ovario. Los datos confirman que  $\Delta 3\_5$  tiene una contribución sustancial a la expresión global del locus *RAD51D* en todos los tipos de muestras analizadas.

## Cuantificación del evento de *splicing* *RAD51D* $\Delta 3\_5$ vs. ( $\Delta 3\_5$ + E3E4) por PCR digital



**Figura 38.** Diagrama de Cajas que muestra los niveles de expresión de  $\Delta 3\_5$  de *RAD51D* (FAM) respecto la combinación del evento alternativo y el mensajero de referencia E3E4 (FAM + VIC) cuantificados por PCR digital. Se muestran los datos agrupados de cada una de las muestras: sangres (n=4), LCLs (n=12), tejido de mama (n=4) y tejido de ovario (n=4). En el caso de los tejidos de mama y ovario la N se refiere a réplicas técnicas. Los diagramas muestran el valor mínimo, percentil 25%, mediana, percentil 75% y valor máximo. Los niveles de  $\Delta 3\_5$  para el tejido de ovario eran superiores que para el resto de tejidos ( $\approx 70\%$  respecto a un  $\approx 40\%$  en sangres, LCLs y mama). Dado que sólo hemos trabajado con una muestra de ovario, no podemos saber si estamos ante un efecto tejido específico o una característica propia de la muestra analizada.

#### 4. Caracterización del *splicing* alternativo en el gen de susceptibilidad *PALB2*

En total se han confirmado 16 eventos de *splicing* alternativo en *PALB2* (Tabla 16), de los cuáles 9 introducen un codón de parada prematuro (PTC-NMD), 6 mantienen la fase de lectura (No Fs) y 1 evento elimina el ATG iniciador de la traducción (Non-coding). Otros eventos no confirmados se describen en la Tabla 17.

Eventos <sup>1</sup>	HGVS <sup>2</sup>	Gencode <sup>3</sup>	Biotipo <sup>4</sup>	Anotación Funcional <sup>4</sup>	Lecturas <sup>5</sup>	RT-PCR/EC
Δ1q <sub>169</sub>	r.-121_48del	No descrito	Donor shift	Non-coding	1	Figura 39
Δ1q <sub>17</sub>	r.32_48del	ENST00000568219	Donor shift	PTC-NMD	1	PCR E1E3*
E1 <sub>600</sub> <sup>##</sup>	-	ENST00000561514	Terminal Modification	No-FS	3,5	n.t.
▼E1B <sup>134</sup>	r.48_49ins48+2518_48+2651	ENST00000567003	Casette	PTC-NMD	5,5	Figura 39
Δ4,5	r.212_2514del	ENST00000565038	Multi-casette	PTC-NMD	10,5#	Figura 40
Δ7p <sub>25</sub>	r.2587_2611del	No descrito	Acceptor shift	PTC-NMD	4,5	Figura 43
Δ7p <sub>10</sub>	r.2587_2596del	No descrito	Acceptor shift	PTC-NMD	12,5	Figura 43, 44
Δ7	r.2587_2748del	No descrito	Casette	No FS	1	Figura 43, 44
Δ8,9	r.2749_2996del	No descrito	Multi-casette	PTC-NMD	1	PCR E7E11
Δ9p <sub>30</sub>	r.2835_2864del	No descrito	Acceptor shift	No FS	1	Figura 45
Δ9	r.2835_2996del	No descrito	Casette	No FS	1	Figura 45
Δ9,10	r.2835_3113del	No descrito	Multi-casette	No FS	6,5#	PCR E7E11
Δ11	r.3114_3201del	No descrito	Casette	PTC-NMD	21,5	Figura 46, 47
Δ11,12	r.3114_3350del	No descrito	Multi-casette	No FS	5,5	Figura 47
▼12p <sup>65</sup>	r.3201_3202ins3202-65_3202-1	No descrito	Acceptor shift	PTC-NMD	4,5	Figura 47
Δ12	r.3202_3350del	ENST00000566069	Casette	PTC-NMD	34,5	Figura 47

**Tabla 16. Eventos de *Splicing* Alternativo en *PALB2* confirmados.** En Materiales y Métodos se detalla el sistema de designación de eventos. <sup>2</sup> Se siguen las reglas publicadas por la *Human Genome Variation Society* (<http://vRNAomen.hgvs.org/>). <sup>3</sup> Se utilizó Ensembl (<http://www.ensembl.org/>) para comprobar si el evento aparece descrito en la compilación 26 de GENCODE. <sup>4</sup> Biotipo y Anotación Funcional según los criterios desarrollados por GENCODE (Mudge et al, 2013). <sup>5</sup> Se muestra el número de lecturas obtenido en el experimento realizado en LCL + puro (media de los valores obtenidos con STAR y TopHat). En el mismo experimento, el número de lecturas (media de STAR y TopHat) correspondiente a cada uno de los 12 intrones de referencia osciló entre 156 y 407 (media 297). La cobertura alcanzada en los experimentos realizados en otras muestras biológicas fue sensiblemente inferior, de tal modo que el valor medio de un intrón de referencia fue 86 en la muestra LCL-puro, 197 en Mama, y 21 en fimbria ovárica. \*Eventos que han sido detectados en los ensayos señalados, pero se detectan como productos de muy baja intensidad. Y sólo se detectan en algunas muestras. Por ello no se visualizan en los ensayos mostrados. #Se destacan eventos con número de lecturas muy discrepantes en STAR y TopHat (por ejemplo, Δ4,5 tiene 1 lecturas en STAR y 20 en TopHat). ## Hablando con propiedad, E1<sup>600</sup> no es un evento de *splicing* alternativo, sino el producto de un origen de transcripción alternativo. No se ha testado (n.t) por RT-PCR/EC, pero lo consideramos un evento confirmado al aparecer descrito en la compilación 26 de GENCODE (ENST00000561514). El % de lecturas corresponde a la relación entre el número de lecturas obtenido para cada evento de *splicing* alternativo, y las lecturas de los transcritos de referencia (representados por la unión entre exones), en el análisis combinado de LCLs + puro, LCLs, mama y Fimbria analizados con los algoritmos STAR y TopHat. \*n.t. hace referencia aquellos eventos que no han sido testado.

Eventos <sup>1</sup>	HGVS <sup>2</sup>	Gencode <sup>3</sup>	Biotipo <sup>4</sup>	Anotación Funcional <sup>4</sup>	Lecturas <sup>5</sup>	RT-PCR/EC
▼E1B <sup>134</sup> +Δ2p <sub>6</sub>	r.48_49ins48+2518_48+2651+r.49_54del	No descrito	mixed	PTC-NMD	-	PCR E1E3*
Δ2	r.49_108del	No descrito	Casette	No Fs	-	Figura 39
Δ2p <sub>6/9</sub>	r.49_54del/ r.49_57del	No descrito	acceptor shift	No FS	-	Figura 39
Δ4	r.212_1684del	No descrito	casette	No FS	-	PCR E3E5
Δ4,5p <sub>24</sub>	r.212_1708del	No descrito	mixed	No FS	-	PCR E3E5
Δ5p <sub>24</sub>	r.1685_1708del	No descrito	acceptor shift	No FS	-	PCR E4E5
Δ5	r.1685_2514del	No descrito	casette	No FS	-	PCR E4E6
Δ6	r.2515_2586del	No descrito	casette	No FS	-	Figura 43
Δ10	r.2997_3113del	No descrito	casette	No FS	-	PCR E7E11
Δ11+▼12p <sup>65</sup>	r.3114_3201del+r.3201_3202ins3202-65_3202-1	No descrito	mixed	PTC-NMD	-	Figura 47
E1 <sub>600</sub> +▼E1B <sup>134</sup>	-	No descrito	Mixed	PTC-NMD	1	PCR E1E3
Δ2_5	r.49_2514del	No descrito	Multi-casette	No FS	8,5	PCR E1E7
▼8p <sup>224</sup>	r.2748_2749ins2749-224_2749-1	No descrito	Acceptor shift	PTC-NMD	1	PCR E7E11
▼11p <sup>401</sup>	r.3113_3114ins3114-401_3114-1	No descrito	Acceptor shift	PTC-NMD	1	PCR E 10E13

**Tabla 17. Eventos de *splicing* alternativo en *PALB2* no confirmados.** Los datos de RNAseq (señalados en amarillo) sugieren la presencia de al menos 4 eventos de *splicing* alternativo no validados por RT-PCR/EC (ninguno ha sido detectado en las RT-PCRs indicadas).<sup>1</sup> Los datos de RT-PCR/EC (señalados en naranja) sugieren la presencia de al menos 9 eventos de *splicing* alternativo no validados por RNAseq. En Materiales y Métodos se detalla el sistema de designación de eventos.<sup>2</sup> Se siguen las reglas publicadas por la *Human Genome Variation Society* (<http://vrnaomen.hgvs.org/>).<sup>3</sup> Se utilizó Ensembl (<http://www.ensembl.org>) para comprobar si el evento aparece descrito en la compilación 26 de GENCODE.<sup>4</sup> Biotipo y Anotación Funcional según los criterios desarrollados por GENCODE (Mudge et al, 2013).<sup>5</sup> Se muestra el número de lecturas obtenido en el experimento realizado en LCL + puro (media de los valores obtenidos con STAR y TopHat). En el mismo experimento, el número de lecturas (media de STAR y TopHat) correspondiente a cada uno de los 12 intrones de referencia osciló entre 156 y 407 (media 297). La cobertura alcanzada en los experimentos realizados en otras muestras biológicas fue sensiblemente inferior, de tal modo que el valor medio de un intrón de referencia fue 86 en la muestra LCL-puro, 197 en Mama, y 21 en fimbria ovárica. \*Eventos que han sido detectados en los ensayos señalados, pero se detectan como productos de muy baja intensidad. Y sólo se detectan en algunas muestras. Por ello no se visualizan en los ensayos mostrados.

Tabla 18. Eventos de *splicing* de *PALB2* detectados en diferentes tejidos

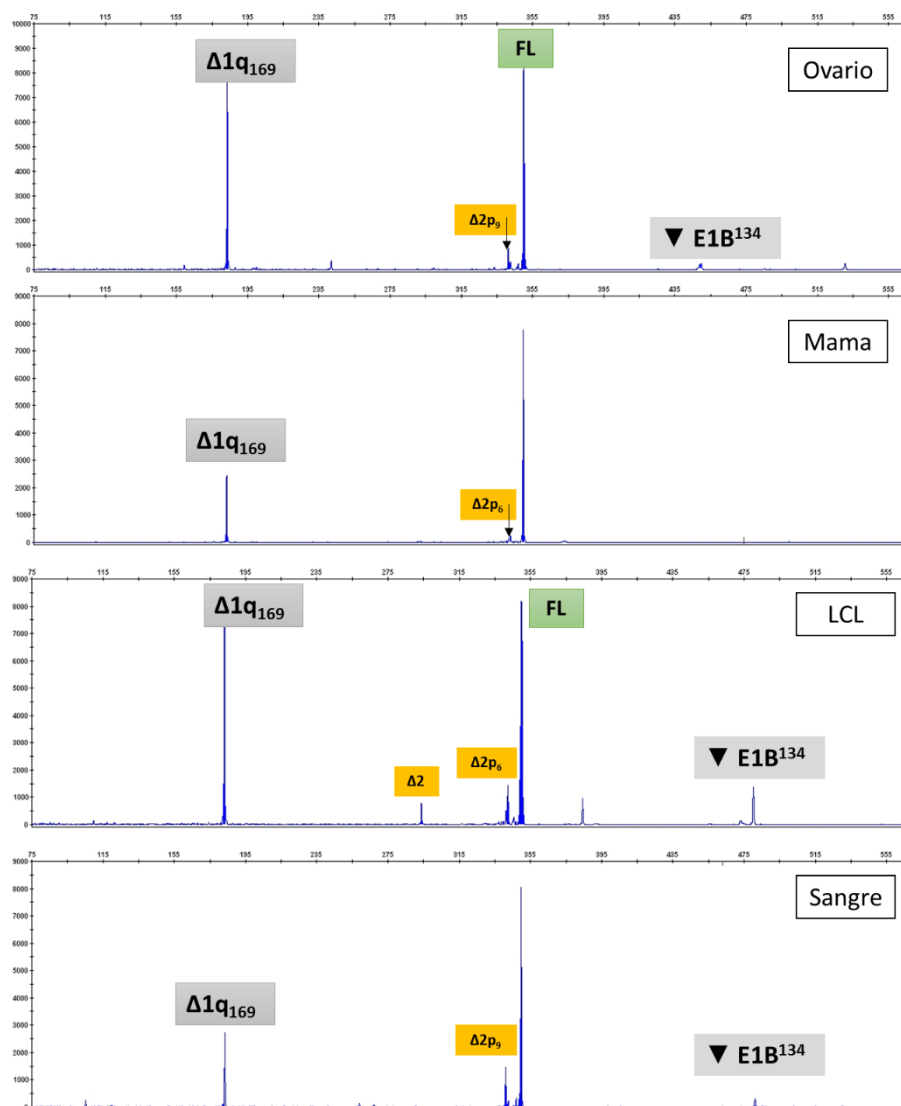
Eventos	LCLs	Sangres	Mama	Ovario
$\Delta 1q_{169}$				
$\Delta 1q_{17}$				
$E1_{600}$		n.t.		n.t.
$E1_{600} + \nabla E1B^{134}$		n.t.		n.t.
$\nabla E1A^{1563}$		n.t.		n.t.
$\nabla E1B^{134}$				
$\Delta 2$				
$\Delta 2p_{6/9}$				
$\Delta 4$				
$\Delta 4,5p_{24}$				
$\Delta 4,5$				
$\Delta 5$				
$\Delta 5p_{24}$				
$\Delta 6$				
$\Delta 7p_{25}$				
$\Delta 7p_{10}$				
$\Delta 7$				
$\nabla 8p^{224}$				
$\Delta 8,9$				
$\Delta 9p_{30}$				
$\Delta 9$				
$\Delta 9,10$				
$\Delta 10$				
$\nabla 11p^{401}$				
$\Delta 11$				
$\Delta 11,12$				
$\nabla 12p^{65}$				
$\Delta 12$				

Tabla 18. Detección de eventos de *splicing* alternativo en *PALB2* (confirmados y no confirmados) estratificados por tipo de muestra biológica. Los recuadros sombreados indican que el evento ha sido detectado por RNAseq y/o RT-PCR/EC, mientras que los recuadros blancos indican que el evento no ha sido detectado. La columna ovario combina los datos obtenidos en ovario mediante RT-PCR/EC, y los datos obtenidos en fimbria ovárica mediante RNAseq. \*n.t. hace referencia aquellos eventos que no han sido testados por RT-PCR/EC. Muchos eventos minoritarios sólo han sido detectados en LCLs. Probablemente, este hecho simplemente refleje la mayor profundidad del análisis realizado en LCLs. Desde un punto de vista clínico, lo más relevante del análisis es la no detección de eventos de *splicing* alternativo específicos de mama u ovario.

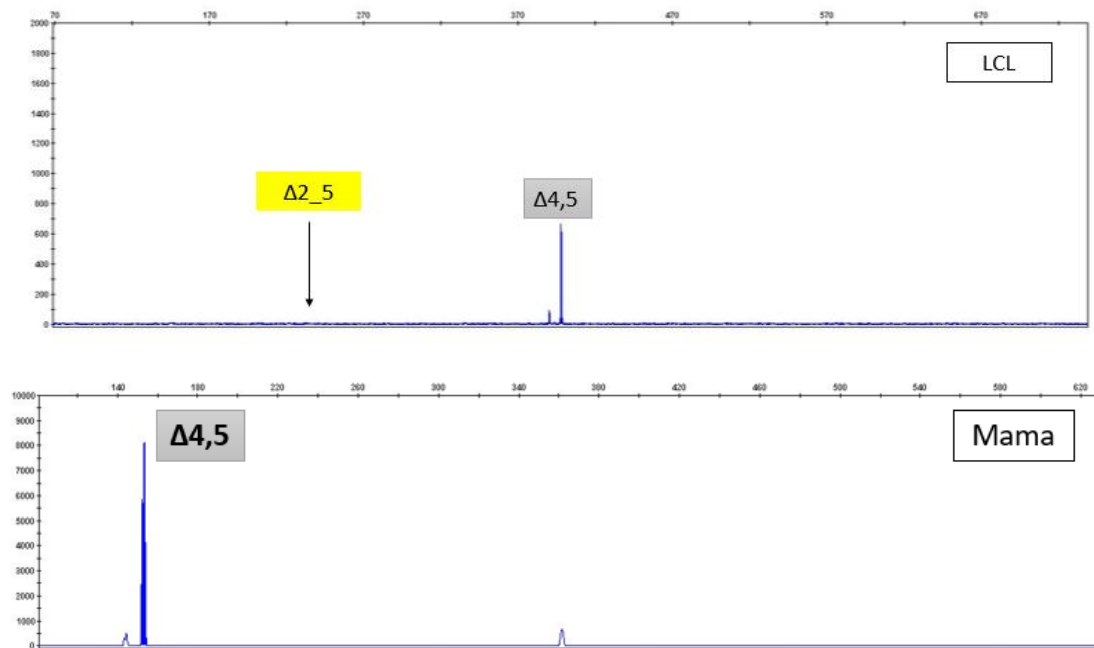


## 4.1. Electroferogramas obtenidos a partir de cada RT-PCR en la caracterización de *PALB2*.

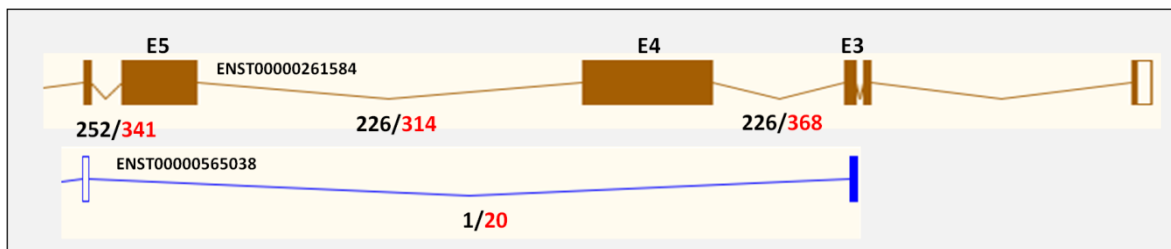
A continuación, se muestran algunos ejemplos representativos de ensayos RT-PCR/EC realizadas en LCLs (N=15), sangres (N=8), mama (N=1; 8 réplicas técnicas) y ovario (N=1; 8 réplicas técnicas), y su relación con los datos de RNAseq. En gris aparecen representados los eventos confirmados (detectados por RNAseq y EC), y en naranja los eventos no confirmados (detectados sólo por EC). En verde se representa el transcrito de referencia.



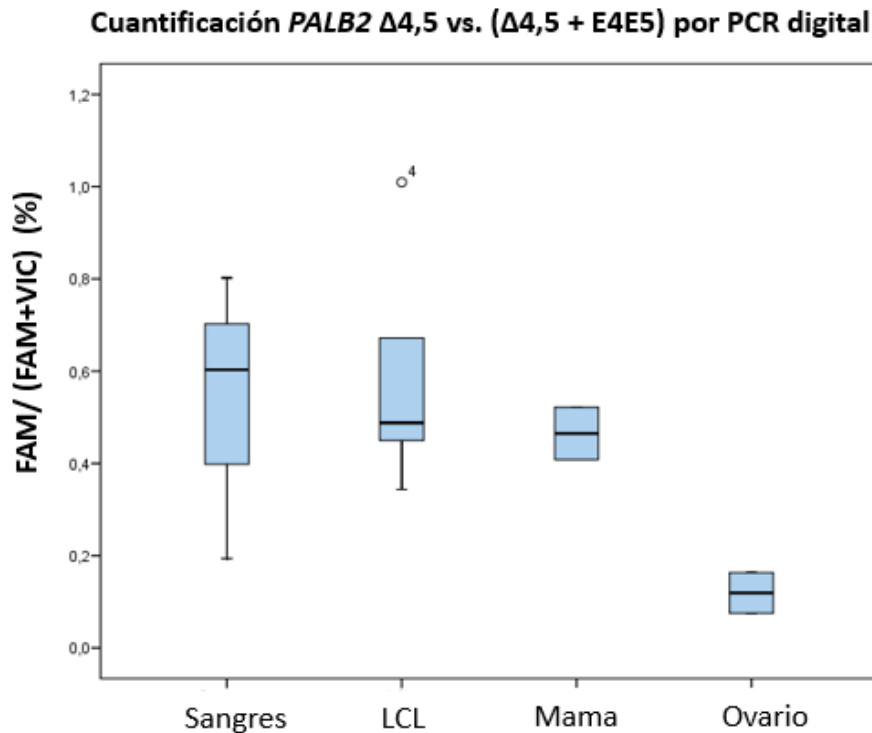
**Figura 39. Ejemplos representativos de ensayo *PALB2* E1-E3 realizados en mama, sangres, LCLs y ovario.** Se detecta el transcrito de referencia (350nt) y los eventos de *splicing* alternativo  $\Delta 1q_{169}$  (183nt),  $\Delta 2$  (293nt),  $\Delta 2p_{6/9}$  (344/341nt), y  $\nabla E1B^{134}$  (480nt). Aunque no ocurra en estos ensayos concretos, los 4 eventos de *splicing* alternativo se han detectado en los cuatro tipos de muestra biológica. Con el signo de interrogación (?) señalamos un potencial evento de *splicing* no anotado (no hemos sabido inferir ninguna anotación en base a la diferencia de tamaño con el transcrito de referencia).



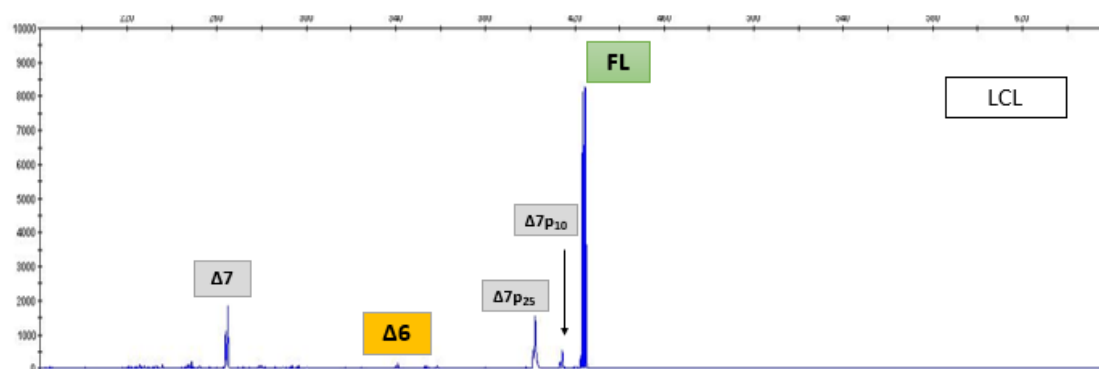
**Figura 40. Ejemplo representativo de ensayos *PALB2* E1E6 y E3E6.** El panel superior muestra un ensayo E1E6 realizado en una LCL, y el panel inferior muestra un ensayo E3E6 realizado en Mama. Los ensayos se realizaron ex profeso para confirmar el evento  $\Delta 4,5$  detectado por RNAseq. En el panel superior se señala la posición donde se esperaba observar la señal correspondiente a *PALB2*  $\Delta 2,5$ , un evento detectado por RNAseq, pero no confirmado por RT-PCR/EC. Los transcritos de referencia de 2684 y 2436 nucleótidos son demasiado grandes para detectarse en nuestras condiciones experimentales. En estas condiciones, los datos de Rt-PCR/EC no nos dicen absolutamente nada sobre la contribución de este evento a la expresión global del gen. Los datos de RNAseq tampoco son muy informativos, tal y como se resume en la Figura 41.



**Figura 41. Análisis por RNAseq de *PALB2*  $\Delta 4,5$ .** Se muestra el transcrito de referencia ENST00000261584 y ENST00000565038 (*PALB2*  $\Delta 4,5$ ). La representación gráfica de los 2 transcritos y los códigos identificativos se han obtenido de Ensembl. La Figura muestra las lecturas de RNAseq correspondientes a los intrones E3E34, E4E5, E5E6 y E3E6. El número de lecturas mostrado se obtuvo en LCL+ puro. En negro se indican las lecturas obtenidas con STAR, y en rojo las lecturas obtenidas con TopHat. De acuerdo con los datos de STAR,  $\Delta 4,5$  es anecdótico (1 lectura), mientras que TopHat sugiere que es un evento cuantitativamente relevante ( $\approx 6\%$  respecto a las lecturas de referencia). Estudios realizados por PCR digital indican que *PALB2*  $\Delta 4,5$  es un evento muy minoritario (Figura 42).



**Figura 42.** Diagrama de Cajas que muestra los niveles de *PALB2*  $\Delta 4,5$  respecto a la combinación del mensajero de referencia y al evento de *splicing* alternativo medidos por PCR digital. Para ello, se diseñó un ensayo TaqMan marcado con FAM que reconoce la unión de los exones 3 y 6 (reconoce  $\Delta 4,5$ ) y un ensayo TaqMan marcado con VIC que reconozca la unión de los exones 4 y 5 (transcrito de referencia de *PALB2* (ver métodos). En el estudio se han incluido muestras de tejido de mama (N=2), tejido de ovario (N=2), LCLs (N=6) y sangres (N=3). En el caso de los tejidos de mama y ovario la N se refiere a réplicas técnicas. Los diagramas muestran el valor mínimo, percentil 25%, mediana, percentil 75% y valor máximo. El círculo blanco indica el valor atípico obtenido en un experimento LCL. Como se observa, *PALB2*  $\Delta 4,5$  representa menos del 1% de la expresión global de *PALB2* en todas las muestras biológicas analizadas.



**Figura 43.** Ejemplo representativo de ensayo *PALB2* ESE8 realizado en una muestra LCL. Se detectan: el transcrito de referencia (FL, 425nt),  $\Delta 7$  (265nt),  $\Delta 6$  (353nt),  $\Delta 7p_{10}$  (415nt) y  $\Delta 7p_{25}$  (402nt). Aunque no ocurra en estos ensayos concretos, los 4 eventos de *splicing* alternativo se han detectado en los cuatro tipos de muestra biológica.

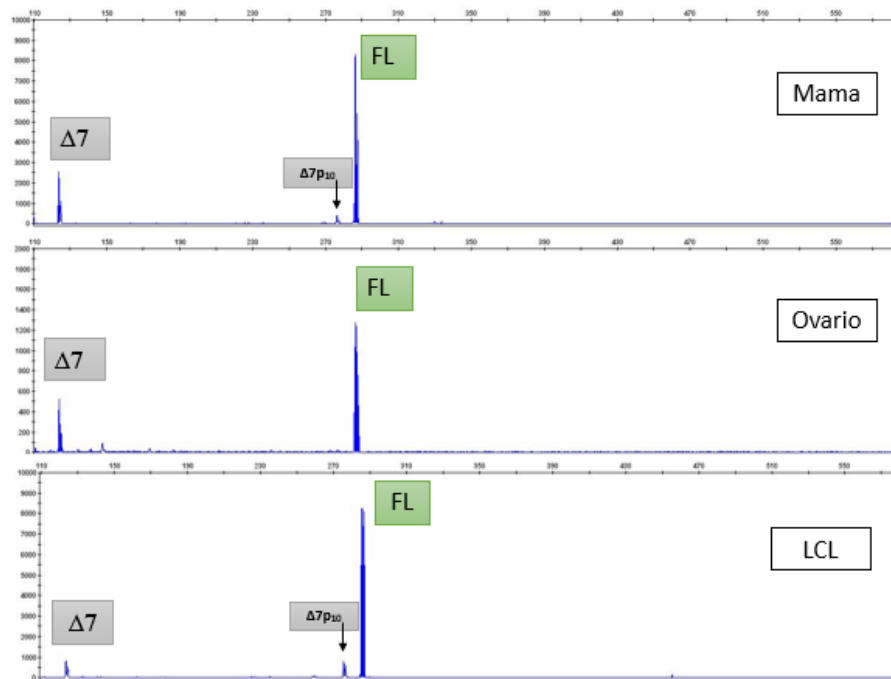


Figura 44. Ejemplo representativo de RT-PCR *PALB2* E6E8 analizados mediante electroforesis capilar en una muestra de LCL, tejido de mama y ovario. Este tipo de experimentos se realizaron en LCLs (N= 11), sangres (N=11), tejido de mama (N=8) y ovario (N=8). En el caso de mama y ovario la N se refiere a réplicas técnicas. Se detectan: el transcrito de referencia (FL, 289),  $\Delta 7$  (123nt), y  $\Delta 7p_{10}$  (275nt). Aunque no ocurra en estos ensayos concretos, los 2 eventos de *splicing* alternativo se han detectado en los cuatro tipos de muestra biológica.

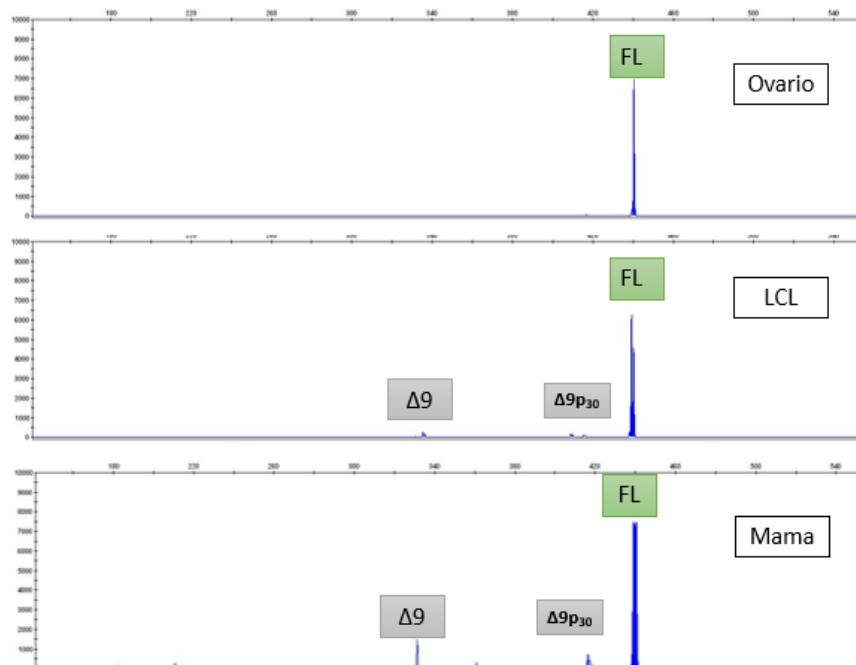
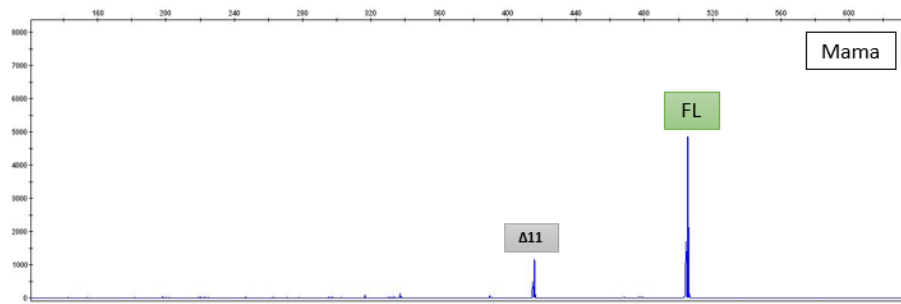
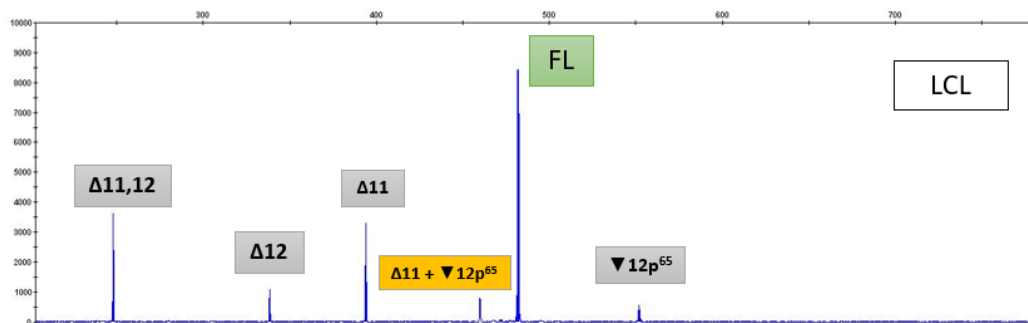


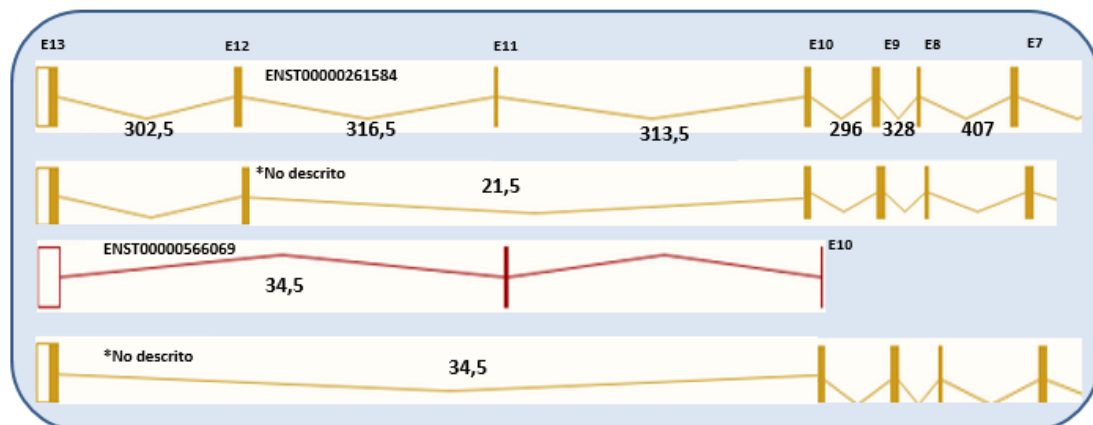
Figura 45. Ejemplo representativo de RT-PCR *PALB2* E7E10 analizados mediante electroforesis capilar en una muestra de LCL, tejido de mama y ovario. Este tipo de experimentos se realizaron en LCLs (N= 10), sangres (N=9), tejido de mama (N=8) y ovario (N=8). En el caso de mama y ovario la N se refiere a réplicas técnicas. Se detectan: el transcrito de referencia (442nt),  $\Delta 9$  (279nt) y  $\Delta 9p_{30}$  (413nt). Aunque no ocurra en estos ensayos concretos, los 2 eventos de *splicing* alternativo se han detectado en los cuatro tipos de muestra biológica.



**Figura 46.** Ejemplo representativo de RT-PCR *PALB2* E9E12 analizados mediante electroforesis capilar en el tejido de mama. Este tipo de experimentos se realizaron en LCLs (N= 10), sangres (N=8), tejido de mama (N=8) y ovario (N=8). En el caso de mama y ovario la N se refiere a réplicas técnicas. Se detectan: el transcrito de referencia (549nt); y  $\Delta 11$  (416nt). Aunque no ocurra en estos ensayos concretos, los 2 eventos de *splicing* alternativo se han detectado en los cuatro tipos de muestra biológica.



**Figura 47.** Ejemplo representativo de RT-PCR *PALB2* E10E13 analizados mediante electroforesis capilar en una muestra de LCL. Este tipo de experimentos también se realizaron en LCLs (N= 12), sangres (N=8), tejido de mama (N=8) y ovario (N=8). En el caso de mama y ovario la N se refiere a réplicas técnicas. Se detectan: el transcrito de referencia (FL, 487nt),  $\Delta 11,12$  (248nt),  $\Delta 12$  (338nt),  $\Delta 11$  (394nt),  $\Delta 11 + \nabla 12p^{65}$  (464nt) y  $\nabla 12p^{65}$  (552nt). Aunque no ocurra en estos ensayos concretos, los 2 eventos de *splicing* alternativo se han detectado en los cuatro tipos de muestra biológica.



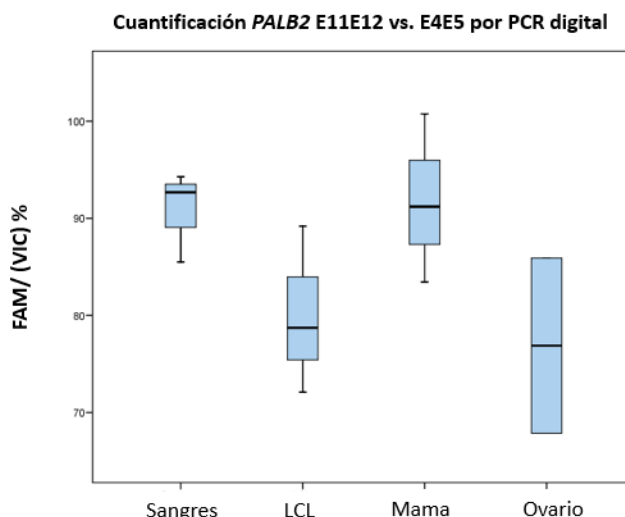
**Figura 48.** Análisis por RNAseq de *PALB2*  $\Delta 11$ ,  $\Delta 12$  y  $\Delta 11,12$ . Se muestra el transcrito de referencia (ENST00000261584), el transcrito  $\Delta 11$  (\*no descrito), el transcrito  $\Delta 12$  (ENST00000566069) y el transcrito  $\Delta 11,12$  (\*no descrito). La representación gráfica de los 4 transcritos y los códigos identificativos se han obtenido de Ensembl. La figura muestra las lecturas de RNAseq correspondientes a los intrones E7E8 (407), E8E9 (328), E9E10 (296), E10E11 (313,5), E12E13 (302,5), E10E12 (21,5), E11E13 (34,5) y E10E13 (34,5). El número de lecturas corresponde a la media de lecturas obtenidas en LCL+ puro con los algoritmos STAR y TopHat.

## 4.2. Contribución de $\Delta 11,12$ a la expresión global de *PALB2*

Los eventos de *splicing* alternativo  $\Delta 11$  y  $\Delta 12$ , así como  $\Delta 11,12$  podrían representar un ejemplo de evento con relevancia clínica, ya que los eventos individuales de *splicing*  $\Delta 11$  y  $\Delta 12$ , no mantienen la fase de lectura, mientras que el evento de *splicing* alternativo  $\Delta 11,12$  sí mantendría la fase de lectura. Por ello, se han diseñado ensayos TaqMan que permitan la cuantificación del evento de *splicing* alternativo  $\Delta 11,12$ .

La cuantificación de este evento de *splicing* alternativo se ha realizado de manera indirecta, mediante la cuantificación de transcritos normales (transcritos que incluyen los exones 10 y 11). La razón principal es que la herramienta de diseño de ensayos TaqMan no admitía la colocación de un cebador entre los exones 9 y 12 de *PALB2*. Por lo que se ha diseñado un ensayo TaqMan marcado con FAM que reconoce la unión de los exones 10 y 11, y de referencia hemos empleado el ensayo TaqMan marcado con VIC que reconoce la unión de los exones 4 y 5.

Los resultados obtenidos por dPCR indican que en todas las muestras analizadas la proporción de mensajeros que expresan transcritos normales (incluyen los exones 10 y 11) es mayor a la de mensajeros que presentan  $\Delta 11,12$ . Es decir, el evento de *splicing* alternativo  $\Delta 11,12$  parece minoritario en todas las formas de tejido analizadas.



**Figura 49. Diagrama de Cajas que muestra los niveles de *PALB2* E11E12 respecto al mensajero de referencia medidos por PCR digital.** Para ello, se diseñó un ensayo TaqMan marcado con FAM que reconoce la unión de los exones 11 y 12; y un ensayo TaqMan marcado con VIC que reconozca la unión de los exones 4 y 5 (transcrito de referencia de *PALB2* (ver métodos). En el estudio se han incluido muestras de LCL (N=3), sangres (N=4), tejido de mama (N=3) y tejido de ovario (N=2). En el caso de los tejidos de mama y ovario la N se refiere a réplicas técnicas. Los diagramas muestran el valor mínimo, percentil 25%, mediana, percentil 75% y valor máximo.

Los resultados obtenidos en este ensayo parecen indicar que los mensajeros de referencia (mensajeros que contienen los exones 11 y 12) pueden representar alrededor de un 70-90% de la expresión global de *RAD51D*, dependiendo del tipo de muestra analizada.

# Discusión





# DISCUSIÓN

---

## 1. Impacto del *splicing* alternativo en la clasificación clínica de variantes genéticas en *BRCA1* y *BRCA2*

### 1.1. Análisis de *splicing* en variantes genéticas sinónimas en *BRCA1* y *BRCA2*

En este trabajo hemos realizado estudios de *splicing* a partir de RNA obtenido de sangre periférica de 13 portadores de mutaciones sinónimas *BRCA1* y 4 en *BRCA2* (ver tablas 8 y 9). Con la excepción de *BRCA1* c.570C>T (p.Thr190=), ninguna de estas variantes se ha clasificado mediante estudio multifactorial. Todas son variantes raras (MAF<1%). De acuerdo a los criterios de clasificación cualitativa del consorcio ENIGMA existentes en el momento de realizar este trabajo, todas estas variantes debían considerarse Clase-3 (relevancia clínica desconocida) y así aparece reflejado en las tablas 8 y 9.

Muy recientemente, el consorcio ENIGMA ha actualizado sus criterios de clasificación de tal modo que ahora los programas de predicción de *splicing* *in silico* tienen un papel fundamental. Si estos programas sugieren un posible efecto a nivel de *splicing*, la variante sinónima debe considerarse Clase-3, de lo contrario debe considerarse Clase-2. En síntesis, ahora el consorcio ENIGMA otorga un valor predictivo negativo a los algoritmos de predicción de *splicing* que no otorgaba en el momento de realizar el trabajo experimental de esta tesis doctoral. En cualquier caso, la hipótesis subyacente es que cualquier variante sinónima, si realmente es patogénica, lo será por alterar el *splicing* (otros mecanismos concebibles, como modificar la estructura de la cromatina y/o los niveles de expresión y/o la estabilidad del RNA, se consideran suficientemente improbables como para poder ignorarse en un contexto de clasificación de variantes con fines diagnósticos). En este escenario, es evidente que una demostración experimental de que una variante sinónima no afecta al *splicing* permite reclasificar dicha variante como Clase-1 (lo que se traduce en indudables beneficios en la consulta de cáncer familiar). Por ello, en nuestro laboratorio consideramos clínicamente relevante analizar el *splicing* en variantes sinónimas, aun sabiendo que el resultado será muy probablemente negativo (y por supuesto, siempre existe la posibilidad de detectar algún efecto positivo). Para demostrar que una variante no afecta al *splicing*, no basta con demostrar que no se observa alteración, además hay que demostrar que el ensayo está capturando la expresión del alelo portador de la variante (el alelo de la variante podría producir mensajeros muy inestables que se degradan y no detectamos, o mensajeros con

retenciones intrónicas muy grandes que no son amplificadas en nuestras condiciones experimentales). Dicho de otro modo, hay que demostrar expresión bi-alélica, y para ello es necesario poder analizar (en nuestro caso mediante secuenciación Sanger) un SNP informativo (heterocigoto en localización exónica). Por definición, los portadores de variantes sinónimas siempre tienen un SNP informativo (la propia variante que estamos testando), simplificando la obtención de un resultado negativo claro. Además, el hecho de poder secuenciar en el cDNA la propia variante nos garantiza que no existe error en la asignación de la muestra, y realmente estamos analizando el *splicing* en RNA de un portador de la variante sinónima de interés. Hay que tener muy presente que el tipo de análisis que hacemos tiene repercusión directa en el asesoramiento y manejo de las familias que atiende la consulta de cáncer familiar de nuestro hospital (y otros hospitales con los que colaboramos) por lo que garantizar que se está analizando la muestra correcta, es esencial.

Los estudios de *BRCA1* c.75C>T (p.Pro25=) mostrados en resultados, y los estudios de *BRCA1* c.5175A>G (p.Glu1725=) y *BRCA2* c.8592C>T (p.Ala2864=) mostrados en el Anexo II son buenos ejemplos de variantes sinónimas que gracias a la combinación de RT-PCR/EC y secuenciación Sanger (se observa expresión bi-alélica balanceada) podemos afirmar que no afectan al *splicing* y se pueden reclasificar como variantes Clase-1. En total, estos estudios nos han permitido reclasificar como Clase-1 un total de 5 variantes sinónimas en *BRCA1* y 4 en *BRCA2*.

Además, los estudios de RT/PCR-EC nos han permitido identificar 2 variantes sinónimas en *BRCA1* (c.528G>A y c.570C>T) que aparentemente (en el mejor de los casos la RT-PCR/EC es una metodología semi-cuantitativa) se asocian con aumentos relativos de los niveles de expresión de la isoforma de *splicing* alternativo *BRCA1*  $\Delta 9,10$  (ver tabla 8). Este dato en sí mismo no permite reclasificar las variantes. Como veremos más adelante (sección 4), ha sido el análisis comparativo por PCR digital del nivel de  $\Delta 9,10$  en portadores de estas variantes y en portadores de c.591C>T (Clase-1 multifactorial) lo que nos ha permitido reclasificar a las dos variantes.

## 1.2. Análisis de *Splicing* en Variantes genéticas *missense* en *BRCA1* y *BRCA2*

Metodológicamente, todo lo que hemos dicho sobre los análisis de *splicing* en variantes sinónimas es igualmente cierto en variantes *missense*: mediante estudios de RT-PCR/EC y secuenciación Sanger en RNA de portadores es relativamente sencillo demostrar que la variante interrogada no afecta al *splicing*. Sin embargo, las implicaciones prácticas a nivel clínico son muy distintas. Es evidente que el hecho de que una variante *missense* no interfiera con el proceso normal de *splicing* no disminuye significativamente la incertidumbre en torno a su significado clínico, y por tanto no permite reclasificar la variante. Esto es así porque lo más probable es que una variante *missense*, si es patogénica, lo sea por sus consecuencias funcionales a nivel de proteína. Sin embargo, un resultado de *splicing* positivo (se detecta alteración) puede permitir demostrar que una variante *missense* es patogénica (Clase-5). De hecho, existen ejemplos relevantes de variantes *missense* en *BRCA1* y *BRCA2* que tienen un efecto deletéreo a nivel de RNA, caso de *BRCA1* p.Arg71Gly (Vega et al, 2001), *BRCA1* p.Arg1495Met (Houdayer et al, 2012), *BRCA2* p.Arg2659Thr (Farrugia et al, 2008) y otros.

Nosotros hemos realizado estudios de *splicing* a partir de RNA obtenido de sangre periférica de 5 portadores de variantes *missense* en *BRCA1* y 9 portadores de variantes *missense* en *BRCA2* (ver tablas 8 y 9). Todas son variantes raras (MAF<1%) sin estudio multifactorial. De acuerdo con los criterios de ENIGMA, todas estas variantes deben considerarse Clase-3 (con la excepción de *BRCA2* p.Arg2336His, ver más adelante). Todos los estudios de *splicing* realizados han sido negativos, con la excepción de los estudios realizados en *BRCA1* p.Pro258Arg y p.Ser316Gly (ambas variantes se asocian con aumento relativo de *BRCA1* Δ9,10) y *BRCA2* p.Arg2336His (causa Δ12,13 y Δ13).

En el Anexo II se muestra un ejemplo representativo de estudio negativo realizado en un portador de la variante *missense* *BRCA1* c.230C>T (p.Thr77Met). El estudio permite descartar que la variante afecte al *splicing*, pero no permite descartar que la variante sea patogénica.

El caso de *BRCA2* c.7007G>A (p.Arg2336His) es distinto. Esta variante está localizada en el extremo 3' de exón 13 y de acuerdo a los criterios de clasificación de ENIGMA, en ausencia de datos complementarios debe considerarse Clase-4 (y no Clase-5) en base al siguiente razonamiento: este tipo de alelos (*sustitución de una G en el extremo de un exón siempre que los 6 primeros nucleótidos del intrón no sea GTRRG*) sólo producirá transcritos alterados, pero existe una probabilidad baja de que estos transcritos alterados no sean deletéreos. Para poder reclasificar este tipo de variantes a Clase-5 es necesario *demostrar*

*experimentalmente que el alelo de la variante sólo produce transcritos deletéreos.* Existen numerosos estudios previos de *splicing* que demuestran que la variante sólo produce transcritos deletéreos (incapaces de codificar proteínas funcionales) y por tanto justifican su reclasificación a Clase-5. Sin embargo, los distintos estudios son discrepantes en cuanto a la naturaleza concreta de la alteración deletérea. En un primer estudio realizado en una LCL portadora de la variante (Thomanssen et al, 2006), un experimento de RT-PCR con cebadores en los exones 12 y 14 detectó únicamente el transcrito  $\Delta 13$  (PTC-NMD). Resultados muy similares fueron obtenidos por Sanz y colaboradores en linfocitos (Sanz et al, 2010). En ambos casos la localización de los cebadores impidió testar si la variante tenía algún efecto en  $\Delta 12,13$ , una isoforma de *BRCA2* por *splicing* alternativo PTC-NMD muy minoritaria (Fackenthal et al, 2016). Un estudio realizado en cultivos primarios de linfocitos tratados con puromicina (Machackova et al, 2008) y RT-PCR con cebadores localizados en los exones 11 y 15 detectó tanto  $\Delta 12,13$  como  $\Delta 13$ . La relevancia del *splicing* alternativo a la hora de elegir el ensayo RT-PCR más adecuado es un tema recurrente de esta tesis doctoral, y una de las razones por la que hemos querido caracterizar el *splicing* alternativo en *PALB2*, *RAD51C* y *RAD51D*. En otro estudio realizado en LCLs, Houdayer y colaboradores detectan  $\Delta 12,13$  (lo describen correctamente como incremento de una isoforma de *splicing* alternativo) pero no  $\Delta 13$  (Houdayer et al, 2012). Nuestro estudio realizado a partir de RNA obtenido de sangre de un portador coincide con Machackova en detectar tanto  $\Delta 12,13$  como  $\Delta 13$  (ver Anexo II). En este caso particular, la discrepancia entre estudios no es particularmente relevante (desde un punto de vista clínico) pues tanto  $\Delta 12,13$  como  $\Delta 13$  son transcritos PTC-NMD, pero en otras ocasiones puede ser fundamental (ver el caso de *BRCA1* c.594-2A>C en sección 4).

### 1.3. Análisis de *Splicing* en Variantes Genéticas Intrónicas en *BRCA1* y *BRCA2*

En este trabajo hemos realizado estudios de *splicing* a partir de RNA de sangre periférica de 7 portadores de variantes intrónicas en *BRCA1* y 4 portadores de variantes intrónicas en *BRCA2* (ver tablas 8 y 9). A diferencia de lo comentado en relación a las variantes exónicas (sinónimas y *missense*), los estudios de *splicing* en portadores de variantes intrónicas tienen la desventaja de que no siempre hay SNP informativos adecuados para poder investigar la expresión bi-alélica. Además, la ausencia de la variante (intrónica) en el cDNA puede plantear dudas en cuanto a la posible confusión de muestras (en nuestro laboratorio hemos tomado como norma, en este tipo de muestras, amplificar por PCR sobre la muestra de cDNA

la región intrónica de interés. Esto es posible porque siguiendo nuestros protocolos de extracción de RNA siempre arrastramos una pequeña proporción de DNA genómico).

La presencia de algún SNP informativo nos ha permitido demostrar que las variantes intrónicas *BRCA1* c.81-13C>G, *BRCA1* c.4675+24G>A, *BRCA2* c.7805+17G>A y *BRCA2* c.7805+23G>A no afectan al *splicing*. Un ejemplo representativo de este tipo de análisis se muestra en el Anexo II (*BRCA2* c.7805+17G>A). De acuerdo con los criterios de ENIGMA, las variantes intrónicas raras (MAF<0.01), en ausencia de información complementaria, deben considerarse Clase-3. De acuerdo a esos mismos criterios, los resultados obtenidos nos permiten reclasificar estas variantes como Clase-1. La variante *BRCA1* c.81-13C>G ya contaba con clasificación multifactorial Clase-1 previa al presente estudio de *splicing* (ver sección 4).

Los análisis por RT/PCR-EC sugieren que la variante *BRCA1* c.4675+11A>G tampoco afecta al *splicing* (Tabla 8), pero debido a la ausencia de SNPs informativos en el portador, no hemos podido confirmar formalmente que estamos capturando la expresión de los dos alelos. Por ello, el estudio realizado no nos permite reclasificar la variante como Clase-1.

Los estudio realizados en la presente tesis doctoral nos han permitido detectar alteraciones de *splicing* en portadores de las variantes *BRCA1* c.594-2A>C, *BRCA1* c.4484+1G>T, *BRCA1* c.5152+4A>T, *BRCA1* c.5333-2A>G, *BRCA2* c.68-7T>A y *BRCA2* c.8488-1G>A (ver tablas 8 y 9).

Las variantes *BRCA1* c.594-2A>C y *BRCA2* c.68-7T>A son casos especiales (existen estudios multifactoriales que demuestran que no se asocian a aumento de riesgo) y se tratarán en la sección 4.

De acuerdo a los análisis de RT-PCR/EC y secuenciación Sanger (ver Tabla 8), la variante *BRCA1* c.4484+1G>T (IVS14+1G>T) genera un transcrito alterado  $\Delta 14$  que no mantiene la fase de lectura (transcrito PTC-NMD). De acuerdo a los criterios del consorcio ENIGMA (ver Introducción), en ausencia de información complementaria, las variantes localizadas en el donador de *splicing* del exón 14 de *BRCA1* deben considerarse Clase-4 (y no Clase-5) en base al siguiente razonamiento: este tipo de alelos (IVS $\pm$ 1,2) sólo producirá transcritos alterados, pero existe una probabilidad baja de que estos transcritos alterados no sean deletéreos. Para poder reclasificar las variantes IVS $\pm$ 1,2 como Clase-5 es necesario demostrar experimentalmente que el alelo de la variante sólo produce transcritos deletéreos. En nuestro estudio sólo hemos detectado transcritos deletéreos  $\Delta 14$  (PTC-NMD) por lo que pensamos que nuestro estudio permite reclasificar la variante como Clase-5.

De acuerdo a los criterios del consorcio ENIGMA (ver Introducción), en ausencia de información complementaria las variantes raras (MAF<0.01) en posiciones intrónicas IVS+3\_IVS-3 (como es el caso de *BRCA1* c.5152+4A>T) deben considerarse Clase-3. Para poder reclasificarlas como Clase-4 es necesario demostrar experimentalmente que la variante genera transcritos deletéreos, y para poder reclasificarla como Clase 5 es necesario demostrar que sólo genera transcritos deletéreos (a diferencia de las variantes IVS±1,2 en este caso no podemos asumir que el alelo de la variante no produzca transcritos de referencia, es necesario demostrarlo experimentalmente). Nuestro análisis mediante RT-PCR/EC y secuenciación Sanger (ver resultados, Figura 16) indica que *BRCA1* c.5152+4A>T (IVS18+4A>T) genera un transcrito deletéreo Δ18. El transcrito *BRCA1* Δ18 mantiene la fase de lectura, pero podemos considerarlo deletéreo (incapaz de codificar una proteína funcional) al codificar parte del dominio BRCT esencial para la función de *BRCA1* (ver introducción). De hecho, algunas de las pocas variantes *missense* inequívocamente patogénicas (Clase-5 multifactorial) que se han descrito en *BRCA1* (p.Arg1699Trp, p.Gly1706Glu, p.Ala1708Glu y p.Ser1715Arg) se localizan en este exón (Lindor et al, 2012). Los datos de RT-PCR/EC y secuenciación Sanger mostrados en la Figura 16 (señal FL y Δ18 aproximadamente al 50% en ambos estudios) sugieren que el alelo c.5152+4T sólo expresa transcritos deletéreos Δ18, pero no lo demuestran formalmente (una demostración formal requeriría un portador con un SNP informativo, o un estudio en minigenes), por lo que pensamos que nuestro estudio permite reclasificar la variante *BRCA1* c.5152+4A>T como Clase-4, pero no como Clase-5. En este contexto es importante señalar que ciertos datos indican que alelos *BRCA1* que sólo expresan un ≈20% de transcritos funcionales podrían conferir *haplosuficiencia* (ver sección 4).

Los estudios de RT-PCR/EC y secuenciación Sanger (Anexo II, Figura 69) indican que *BRCA1* c.5333-2A>G (IVS21-2A>G) sólo genera transcritos deletéreos Δ22 (PTC-NMD) y probablemente transcritos también deletéreos Δ22,23 (PTC-NMD) en muy baja proporción (Δ22,23 se han descrito previamente en controles como isoforma por *splicing* alternativo muy minoritaria, Colombo et al, 2014). Como ya se ha comentado en relación a *BRCA1* c.4484+1G>T, este tipo de resultados nos permite reclasificar la variante como Clase-5.

La variante *BRCA2* c.8488-1G>A (IVS19-1G>A) es particularmente interesante porque se ha descrito en homocigosis en un paciente con Anemia de Fanconi subtipo D1 (Howlet et al, 2002). En ese trabajo se demostró mediante RT-PCR + Sanger que la variante causaba Δ20p<sub>12</sub>. Existen al menos otros 3 estudios previos al nuestro que analizan el efecto a nivel de *splicing* de la variante *BRCA2* c.8488-1G>A (IVS19-1G>A). Un estudio realizado en un minigen que

incluía los exones 19 y 20 de *BRCA2* (Acedo et al, 2012) sugirió que la variante generaba principalmente transcritos deletéreos  $\Delta 20$  (PTC-NMD), pero también de forma minoritaria la retención del intrón 19 (PTC-NMD), transcritos  $\Delta 20p_{12}$  y de forma muy minoritaria transcritos  $\Delta 19,20$  (PTC-NMD). En el mismo trabajo, un análisis realizado directamente en linfocitos de un portador detectó mayoritariamente transcritos  $\Delta 20p_{12}$ , de forma minoritaria retención del intrón 19, y no detectó  $\Delta 20$ . Un estudio posterior realizado a partir de RNA extraído de sangre de portadores (RT-PCR/Agarosa con cebadores en exones 18 y 21) concluyó que la variante produce un transcrito principal  $\Delta 20p_{12}$ , y un transcrito minoritario  $\Delta 20$ , pero no detectó retención del intrón 19 (Santos et al, 2014). Otro estudio realizado en un minigen que incluía los exones 19 a 27 de *BRCA2* (Acedo et al, 2015) sugiere que el alelo de la variante es capaz de producir hasta 5 transcritos distintos, que ordenados en función de los niveles de expresión de mayor a menor serían: retención del intrón 19,  $\Delta 20$ ,  $\Delta 20p_{12}$ , un transcrito no bien caracterizado que implica la inserción de 23 nucleótidos, y cantidades residuales del transcrito de referencia.

Nuestros estudios de RT/PCR-EC y secuenciación Sanger (Anexo II, Figura 78) realizados en un portador de la variante *BRCA2* c.8488-1G>A indican que el alelo de la variante produce mayoritariamente  $\Delta 20p_{12}$ , pero también una pequeña proporción de  $\Delta 20$  y retención del intrón 19.

La base de datos BRCashare clasifica la variante como Clase-5 aparentemente en base a los estudios de *splicing* (algo en lo que no estamos de acuerdo). La base de datos ClinVar la considera Clase-4/Clase-5. Nosotros pensamos que en base a los datos de *splicing* existentes la variante debe considerarse Clase-3. Por otro lado, la ya comentada identificación de un paciente FA-D1 homocigoto para esta variante (Howlet et al, 2002) sugiere, en nuestra opinión, que *BRCA2* c.8488-1A es un alelo hipomorfo, algo que podría ser explicado por la expresión de la proteína *BRCA2* p.(Thr2830\_Lys2833del).

### 1.4. Análisis de *splicing* en variantes Clase-1 multifactorial

Hemos realizado diversos estudios de *splicing* en 3 variantes genéticas en *BRCA1* (c.81-13C>G, c.591C>T, c.594-2A>C) y 2 variantes genéticas en *BRCA2* (c.68-7T>A, c.223G>C) que son Clase-1 multifactorial (Lindor et al, 2012, de la Hoya et al, 2016; en el caso de c.68-7T>A el estudio multifactorial está en proceso de publicación). Como se verá, la posible utilidad de



estos estudios radica en identificar y caracterizar con la mayor precisión posibles alteraciones de *splicing* que sabemos que no se asocian con un incremento de riesgo. En algunos casos, esta información nos permitirá deducir reglas de clasificación aplicables a otras variantes para las que no existen datos multifactoriales.

Como ya se ha comentado, los estudios de RT-PCR/EC y secuenciación Sanger realizados en un portador de la variante *BRCA1* c.81-13C>G no han identificado ningún tipo de alteración de *splicing*, ni cuantitativa (nuevos transcritos), ni cualitativa (variaciones en los niveles de expresión relativo de algunas isoformas). Por ello, el estudio de esta variante no nos ha aportado ninguna información útil.

Los estudios de RT/PCR-EC y secuenciación Sanger realizados por nuestro grupo de investigación hace unos años (Dosil et al, 2010) más los análisis realizados por PCR digital en la presente tesis doctoral (resultados, Figura 15) indican que la variante c.591C>T expresa fundamentalmente transcritos  $\Delta 9,10$  (también expresa una pequeñísima proporción de transcritos deletéreos  $\Delta 9$ ) y no expresa niveles detectables de mensajero de referencia. *BRCA1*  $\Delta 9,10$  es una de las isoformas por *splicing* alternativo de *BRCA1* más abundantes (Colombo et al, 2014), representando aproximadamente el 18% de los transcritos *BRCA1* en controles (resultados, Figura 15). En nuestra opinión, el dato de que el alelo c.591T no es patogénico (Clase-1) demuestra que *BRCA1*  $\Delta 9,10$  debe codificar una proteína *BRCA1* p.(Gly183\_Lys223del) funcionalmente equivalente a la proteína *BRCA1* de referencia. Por esta razón, consideramos que cualquier variante sinónima o intrónica cuyo único efecto sobre el *splicing* sea aumentar la contribución de *BRCA1*  $\Delta 9,10$  a la expresión global de *BRCA1* puede considerarse Clase-1. Este es el caso de las variantes sinónimas c.528 G>A (p.Thr176=) y c.570C>T (p.Thr190=) (Figura 67). El análisis de *splicing* de la variante c.594-2A>C (IVS9-2A>C) ha sido particularmente revelador. Durante muchos años esta variante se ha considerado patogénica al estar situada en una zona consenso de *splicing*. En el año 2005 se publicó un estudio *in vitro* de *splicing* en linfocitos de portadores demostrando que la variante provocaba la pérdida del exón 10 de *BRCA1* en el mensajero maduro (Tesoriero et al., 2005). Esta alteración provoca un desplazamiento del marco de lectura e introduce un codón de parada prematura, lo que parecía justificar plenamente el carácter patogénico atribuido a la variante. Sin embargo, en el estudio de Tesoriero la localización de los cebadores no permitía capturar información de la isoforma *BRCA1*  $\Delta 9,10$  (de nuevo surge el tema, recurrente en esta tesis doctoral, de la importancia del *splicing* alternativo para seleccionar correctamente la localización de los cebadores en un experimento de RT-PCR).

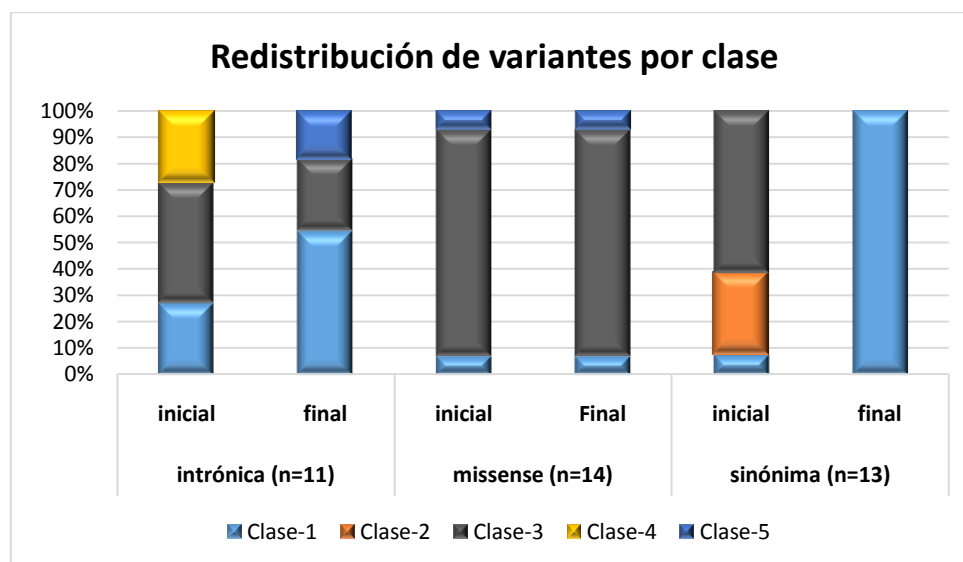
Los estudios de *splicing* que mostramos en el presente trabajo (Figuras 14 y 15) demuestran de forma convincente que la variante c.594-2A>C (IVS9-2A>C), tal y como había mostrado Tesoriero en 2005, altera el *splicing* eliminando el exón 10 del mensajero maduro de *BRCA1*. Sin embargo, esa misma alteración apenas altera los niveles normales de *BRCA1*  $\Delta 9,10$ . La conclusión inevitable del estudio es que los niveles normales de *BRCA1*  $\Delta 9,10$  (aproximadamente un 20% de la expresión global de *BRCA1*) son suficientes para conferir haplosuficiencia.

Es precisamente el estudio combinado (multifactorial y de *splicing*) de las variantes c.591C>T y c.594-2A>C el que justifica que el consorcio ENIGMA, en sus reglas de clasificación cualitativa, considere las variantes en los donadores y aceptores de *splicing* de los exones 9 y 10 de *BRCA1* excepcionales (Introducción Tabla 4). Muy probablemente, todas estas variantes causarán *skipping* de los exones 9 ó 10 (ambas alteraciones PTC-NMD), pero no van a disminuir los niveles normales de *BRCA1*  $\Delta 9,10$  (en todo caso aumentarlos) por lo que es casi seguro que confieran haplosuficiencia. Además, los resultados del análisis de c.594-2A>C permite hacer otras predicciones que aún no se han concretado en reglas de clasificación específicas: 1) es muy probable que los codones de parada prematuros en los exones 9 y 10 de *BRCA1* tampoco sean patogénicos, y 2) es muy probable que cualquier variante genética (por ejemplo en el promotor) que se asocie con una reducción de los niveles de expresión de *BRCA1* de hasta el 80% tampoco sea patogénica.

El análisis mediante RT-PCR/EC mostrado en el Anexo II (Figura 70) indica que la variante *BRCA2* c.68-7T>A se asocia con una alteración cuantitativa del *splicing* (incremento de la isoforma  $\Delta 3$ ), confirmando estudios previos (Vreeswijk MP et al, 2009, Sanz et al, 2010, Houdayer C et al, 2012). La isoforma por *splicing* alternativo  $\Delta 3$  es una de las más abundantes de *BRCA2* (Fackenthal et al, 2016). Esta isoforma mantiene la fase de lectura, aunque codifica parte del dominio de interacción con *PALB2* (ver introducción). Está bien establecido el hecho de que los alelos de *BRCA2* que únicamente expresan transcritos  $\Delta 3$  son patogénicos (Muller et al, 2011), indicando que *BRCA2*  $\Delta 3$  (a diferencia de *BRCA1*  $\Delta 9,10$ ) no es capaz de conferir haplo-suficiencia. Sin embargo, debe existir un nivel máximo tolerable de transcritos *BRCA2*  $\Delta 3$  que esté muy por encima de los niveles observados en controles (o dicho de otro modo, un nivel mínimo tolerable de transcritos *BRCA2*  $\nabla E3$  que esté por debajo del observado en controles), como demuestra el hecho de que *BRCA2* c.68-7T>A no se asocie con un incremento en el riesgo de desarrollar cáncer (estudio multifactorial pendientes de publicación). Por ello, nos pareció particularmente relevante medir con precisión (PCR digital) los niveles de *BRCA2*

$\Delta 3$  en portadores de la variante *BRCA2* c.68-7T>A. Los resultados mostrados en el Anexo II (Figura 71) indican que el nivel de *BRCA2*  $\Delta 3$  aumenta  $\approx 4x$  en portadores (lo que equivale a una reducción modesta, en torno al 10%, de los niveles de *BRCA2*  $\nabla 3$ ). Otros grupos del consorcio ENIGMA han llegado a conclusiones similares utilizando otras metodologías de medición como qPCR y RNAseq (datos pendientes de publicación). A partir del estudio combinado (multifactorial y de *splicing*) de la variante *BRCA2* c.68-7T>A concluimos que cualquier variante sinónima o intrónica en *BRCA2* cuya única alteración de *splicing* sea un incremento en los niveles de expresión de *BRCA2*  $\Delta 3$  igual o inferior al observado en *BRCA2* c.68-7T>A se puede considerar Clase-1.

Realizamos un análisis de *splicing* muy similar en un portador de la variante *BRCA2* c.223G>C (p.Ala75Pro), previamente clasificada como Clase-1 multifactorial (Lindor et al, 2012). El objetivo del análisis era simple: si la variante (localizada en el exón 3 de *BRCA2*) se asociaba con un incremento de los niveles de *BRCA2*  $\Delta 3$  superiores a los observados en *BRCA2* c.68-7T>A, estaríamos aumentando el rango de niveles de *BRCA2*  $\Delta 3$  tolerables. Sin embargo, tanto los datos de RT-PCR/EC (Anexo II, Figura 72), como los datos de PCR digital (Anexo II, Figura 73) indicaron que la variante no afecta a los niveles de *BRCA2*  $\Delta 3$ , por lo que este estudio no nos aportó información útil.



**Figura 50. Proporción de variantes anotadas de cada clase, según el tipo de variante analizada para los locus *BRCA*.** En el gráfico se muestra el porcentaje de variantes anotadas por clase inicial (antes del estudio) y final (después del estudio de *splicing*). Se han analizado un total de 11 variantes intrónicas, 14 missense y 13 sinónimas.

Sorprendentemente, la existencia de *splicing* alternativo tiende a ser ignorada en tanto en los estudios de investigación en susceptibilidad genética al cáncer de mama y ovario, como en su traslación a la clínica. La presente tesis doctoral defiende que esto es un error (en particular, cuando los resultados de la investigación se trasladan a la clínica), y postula que una caracterización detallada del *splicing* alternativo de un gen, aporta información esencial para la correcta interpretación clínica de las variantes genéticas identificadas en dicho gen.

## **2. IMPACTO DEL *SPLICING* ALTERNATIVO EN LA CLASIFICACIÓN CLÍNICA DE VARIANTES GENÉTICAS EN *RAD51C*, *RAD51D* y *PALB2***

### **2.1. Clasificación clínica inicial de variantes genéticas en *RAD51C*, *RAD51D* y *PALB2***

En la mayor parte de los estudios de investigación realizados para identificar nuevos genes de susceptibilidad al cáncer de mama y ovario (y/o para cuantificar con precisión el riesgo atribuible a genes ya establecidos) se asume que todas las variantes germinales en sitios consenso de *splicing* (IVS±1,2), y todas las variantes que introducen un codón de parada prematuro (PTCs), con la posible excepción de PTCs localizadas en el último exón codificante, son variantes de pérdida de función (y por tanto patogénicas), sin necesidad de realizar estudios funcionales y/o estimaciones directas de riesgo que lo avalen (Goldgar DE et al, 2004; Easton et al, 2015).

Nosotros pensamos que lo anterior es cierto para la mayoría de las variantes genéticas, y por tanto es una generalización útil en muchos estudios de investigación. Sin embargo, también pensamos que cuando los estudios genéticos se trasladan del campo de la investigación al campo del diagnóstico clínico, se debe ser más conservador y tratar ciertas variantes IVS±1,2 y PTC con especial cautela (Clase-3).

¿Qué variantes IVS±1,2 y PTC deben tratarse con especial cautela? Una de las tesis centrales del presente trabajo es que los eventos de *splicing* alternativo (en particular, si codifican proteínas potencialmente funcionales) nos ayuda a identificar este tipo de variantes, tal y como se argumenta a continuación:

i) Una variante IVS±1,2 alterará con toda probabilidad el *splicing*, pero no es menos cierto que el resultado de dicha alteración no tiene porqué ser necesariamente patogénico (podría por ejemplo resultar en una pequeña alteración en el mensajero que mantenga la fase de lectura y codifique una proteína funcional). En este contexto, nos parece de particular relevancia la existencia de eventos de *splicing* alternativo acceptor shift y donor shift que

mantengan la fase de lectura, pues estos eventos identifican exones que tienen al menos dos aceptores y/o donadores de *splicing* funcionales. Esto sugiere a su vez un mecanismo obvio para que las variantes IVS $\pm$ 1,2 produzcan transcritos que mantengan la fase de lectura (las variantes destruirán el aceptor/donador de referencia, pero no necesariamente el aceptor/donador alternativo). Estos transcritos pueden por supuesto codificar proteínas no funcionales, pero también proteínas parcial o completamente funcionales. En el mismo sentido, también pueden requerir especial consideración variantes IVS $\pm$ 1,2 en exones que, aun careciendo de aceptores y/o donadores de *splicing* alternativos, mantengan la fase de lectura y no codifiquen dominios esenciales para la correcta función de la proteína. En síntesis, estos son los argumentos utilizados por el consorcio ENIGMA para considerar Clase-3 las variantes localizadas en el aceptor de *splicing* de los exones 8, 13 y 14 de *BRCA1*, y en el aceptor y donador del exón 12 de *BRCA2* (ver introducción, Tabla 4), y nos parecen argumentos igualmente válidos para aplicarlos a variantes IVS $\pm$ 1,2 en *RAD51C*, *RAD51D* y *PALB2*.

ii) Una variante PTC inactivará con toda probabilidad todos los transcritos donde esté presente. Sin embargo, es posible que la variante PTC se localice en una región de la secuencia codificante no constitutiva, y que las isoformas de *splicing* alternativo que no contengan dicha región (y por lo tanto no contengan la variante PTC) proporcionen haplosuficiencia (lo que implica a su vez que proporcionen un nivel de expresión suficiente de una proteína funcional). Eso es precisamente lo que sugiere el estudio de la variante *BRCA1* IVS9-2 que presentamos en esta tesis doctoral. Aunque se trata de una variante IVS-2, causa pérdida del exón 10 de *BRCA1*, lo que a todo los efectos es equivalente a una variante PTC en dicho exón. Sin embargo, la variante no es patogénica porque los niveles normales de *BRCA1*  $\Delta$ 9,10 (alrededor de un 20% de la expresión global de *BRCA1*) confieren haplosuficiencia. Del mismo modo, es teóricamente posible que existan isoformas por *splicing* alternativo que confieran haplosuficiencia y “protejan” del efecto deletéreo de ciertas variantes PTC en *RAD51C*, *RAD51D* y/o *PALB2*.

### 2.1.1. *RAD51C*

En el presente trabajo hemos identificado en *RAD51C* un total de 9 eventos de *splicing* alternativo que mantienen la fase de lectura ( $\Delta$ 2p<sub>3</sub>,  $\Delta$ 3q<sub>114</sub>,  $\Delta$ 3q<sub>114</sub>-4p<sub>81</sub>,  $\Delta$ 4p<sub>81</sub>,  $\Delta$ 4p<sub>111</sub>,  $\Delta$ 5,  $\Delta$ 6\_8,  $\nabla$ 8p<sup>3</sup> y  $\Delta$ 9p<sub>6</sub>). Todos estos eventos están confirmados (Tabla 10), aunque son muy minoritarios (de acuerdo a los datos de RNAseq, las lecturas correspondientes a  $\Delta$ 3q<sub>114</sub>- $\Delta$ 4p<sub>81</sub>

representan aproximadamente el 3% de las lecturas correspondientes al transcrito de referencia, y la contribución de los otros eventos a la expresión global es incluso menor).

A pesar de mantener la fase de lectura, los transcritos  $\Delta 3q_{114\_4p_{81}}$ ,  $\Delta 4p_{81}$ ,  $\Delta 4p_{111}$ ,  $\Delta 5$  y  $\Delta 6\_8$  codifican proteínas que, de expresarse, carecerían de una de las 8 láminas- $\beta$  que se consideran esenciales para la integridad estructural del dominio C-terminal (Miller KA et al, 2004), por lo que consideramos muy improbable que fuesen funcionales. El dominio C-terminal (residuos 96-376) es crítico para la función de *RAD51C*, mediando su interacción con RAD51B, RAD51D y XRCC3, además de contener la actividad ATPasa y la señal de localización nuclear (Miller KA et al, 2004). En resumen, consideramos muy improbable que  $\Delta 3q_{114\_4p_{81}}$ ,  $\Delta 4p_{81}$ ,  $\Delta 4p_{111}$ ,  $\Delta 5$  y  $\Delta 6\_8$  tengan alguna relevancia en la clasificación clínica de variantes genéticas en *RAD51C*.

Sin embargo, pensamos que  $\Delta 2p_3$ ,  $\nabla 8p^3$  y  $\Delta 9p_6$  si pueden ser relevantes. Consideramos posible que las variantes en los aceptores de *splicing* de los exones 2, 8 y 9 no produzcan (o no sólo produzcan) el *skipping* del exón correspondiente (en todos los casos alteraciones PTC-NMD) sino mensajeros que mantengan la fase de lectura y, potencialmente, puedan codificar proteínas funcionales. En concreto transcritos  $\Delta 2p_3$  (r.146\_148del),  $\nabla 8p^3$  (r.965\_966ins966-1\_966-3) y  $\Delta 9p_6$  (r.1027\_1032del) que predicen la síntesis de proteínas con secuencias primarias muy similares a la proteína de referencia (respectivamente p.Glu49del, p.Glu321\_Arg322insSer y p.Pro343\_Glu344del). El residuo 49 se localiza en el dominio N-terminal (residuos 1-66) necesario para la interacción de *RAD51C* con el DNA (UniProtKB-O43502), pero no es evidente que la pérdida de un único residuo afecte a esta actividad. Las alteraciones p.Gln321\_Arg322insSer y p.Pro343\_Glu344del se localizan en el dominio C-terminal, pero no eliminan ninguna región considerada crítica (láminas- $\beta$ , dominios Walker-A y B, y señal de localización nuclear). En resumen, no es obvio que estas alteraciones sean deletéreas por lo que recomendamos que, en ausencia de estudios *in vitro* de *splicing*, las variantes genéticas en los aceptores de los exones 2, 8 y 9 de *RAD51C* se consideren Clase-3.

Ya hemos comentado que  $\Delta 2p_3$ ,  $\nabla 8p^3$  y  $\Delta 9p_6$  representan una fracción muy pequeña de la expresión global de *RAD51C*, por lo que consideramos muy improbable que, aun codificando proteínas funcionales, confieran haplosuficiencia y pueda “rescatar” variantes PTC. Por ello, no vemos razones para no considerar Clase-5 todos los PTCs identificados en los exones 1 a 8 de *RAD51C*. Recomendamos ser más conservadores con las variantes PTCs

identificadas en el último exón (exón 9) del gen, aunque esto nada tiene que ver con el *splicing* alternativo.

### 2.1.2. RAD51D

En *RAD51D* hemos identificado 11 eventos de *splicing* alternativo que eliminan parte de la secuencia codificante de referencia manteniendo la fase de lectura ( $\Delta 2p_{21}$ ,  $\Delta(2,3)$ ,  $\Delta(3,4)$ ,  $\Delta 3_5$ ,  $\Delta 3_6$ ,  $\Delta 5p_{99}$ ,  $\Delta 5$ ,  $\Delta 6$ ,  $\Delta 7$ ,  $\Delta(7,8)$  y  $\Delta 7_9p_{108}$ ), aunque no todos los eventos están confirmados (Tabla 13). Con la excepción de  $\Delta 5$  y  $\Delta 3_5$ , todos los eventos representan un % muy bajo de la expresión global (estimado a partir de los datos de RNAseq y/o RT-PCR/EC). Por el contrario, las lecturas que soportan el evento  $\Delta 5$  representan aproximadamente el 8% de las lecturas de referencia, y las lecturas que soportan  $\Delta 3_5$  aproximadamente el 40% (la gran contribución de  $\Delta 3_5$  a la expresión global de *RAD51D* se ha confirmado por PCR digital, ver Figura 38). Sin embargo, consideramos muy improbable que transcritos que carecen del exón 5 (residuos 116-160), o de los exones 3, 4 y 5 (residuos 48-160) codifiquen para proteínas funcionales. Los exones 3 y 4 codifican gran parte del dominio N-terminal (implicado en la unión a XRCC2 y a DNA de hebra sencilla), la región *linker*, y regiones que se presumen esenciales para la funcionalidad del dominio C-terminal (interacción con *RAD51C* y unión e hidrólisis de ATP). En concreto, los exones 3 y 4 codifican la primera de las 8 láminas- $\beta$  esenciales en la estructuración del dominio, y el motivo Walker A, esencial para la unión e hidrólisis de ATP. El exón 5 codifica la segunda de las dos láminas- $\beta$  esenciales en la estructuración del dominio C-terminal (Miller A et al., 2004; Kim and Choi 2011; Suwaki et al, 2011; UniProtKB - O75771). Por estas razones, consideramos improbable que  $\Delta 5$  y  $\Delta 3_5$  sean relevantes para la clasificación clínica de variantes genéticas identificadas en *RAD51D*. Las mismas consideraciones estructurales nos llevan a pensar que  $\Delta(2,3)$ ,  $\Delta(3,4)$ ,  $\Delta 3_6$ ,  $\Delta 5p_{99}$ ,  $\Delta 6$ ,  $\Delta 7$ ,  $\Delta 7,8$  y  $\Delta 7_9p_{108}$  no codifican proteínas *RAD51D* funcionales (ni ser relevantes para la clasificación clínica de variantes).

Sin embargo, pensamos que  $\Delta 2p_{21}$  si puede ser relevante. Consideramos posible que las variantes en el aceptor de *splicing* del exón 2 no produzcan (o no sólo produzca) el *skipping* del exón 2 (alteración PTC-NMD), sino transcritos r.83\_103del que codificarían para una proteína p.(Val28\_Ala34) que carece de 7 residuos en el extremo N-terminal de la proteína. Ya hemos comentado que el extremo N-terminal de *RAD51D* (residuos 1-83) es funcionalmente relevante, pero desconocemos el efecto concreto que podría tener la pérdida de estos 7 residuos, y no podemos asumir que causasen pérdida de función. Por esta razón pensamos que, en ausencia de estudios *in vitro* de *splicing*, las variantes *RAD51D* localizadas en el aceptor de *splicing* del exón 2 deben considerarse Clase-3. Probablemente,  $\Delta 2p_{21}$  representa una



fracción muy pequeña de la expresión global de *RAD51D* (no se ha detectado por RNAseq) por lo que consideramos muy improbable que, aun codificando una proteína funcional, los niveles normales de expresión de este evento confieran haplosuficiencia y pueda “rescatar” variantes PTC.

No obstante, el análisis de posibles PTCs no patogénicos en *RAD51D* se complica por el hecho de que hay dos transcritos considerados frecuentemente de referencia que contienen exones 3 alternativos. Nuestros resultados indican que el exón 3 de 119nt que corresponde al transcrito ENST0000345365 (codifica para una proteína RAD51D de 328 aa) se expresa en mayor medida que el exón 3 de 179nt del transcrito ENST0000590016 (codifica una proteína RAD51D de 348 aa), y por ello utilizamos ENST0000345365 como transcrito de referencia. Sin embargo, los niveles de expresión del exón 3 de 179nt son apreciables. De acuerdo a los datos de RNAseq, representan  $\approx 20\%$  de las lecturas correspondientes al exón 3 de 119 nucleótidos. Los datos obtenidos por PCR digital son concordantes (ver Figura 29). Desconocemos si la función biológica de la proteína RAD51D<sup>348aa</sup> es equivalente, complementaria o antagónica a la de RAD51D<sup>328aa</sup>. En cualquier caso, consideramos prudente tratar con especial precaución las variantes PTC localizadas en cualquiera de los dos exones 3 alternativos de *RAD51D* y considerarlas, en ausencia de estimaciones directas de riesgo, variantes Clase-3 (por idéntica razón, recomendamos considerar Clase-3 las variantes en los donadores y aceptores de *splicing* de los dos exones 3). También recomendamos ser conservadores en la clasificación clínica de variantes PTCs identificadas en el último exón codificante (exón 10) del gen, aunque esto nada tenga que ver con el *splicing* alternativo.

### 2.1.3. *PALB2*

En *PALB2* hemos identificado hasta 13 eventos de *splicing* alternativo que eliminan parte de la secuencia codificante manteniendo la fase de lectura ( $\Delta 2p_{6/9}$ ,  $\Delta 2$ ,  $\Delta 2\_5$ ,  $\Delta 4$ ,  $\Delta(4,5p_{24})$ ,  $\Delta 5p_{24}$ ,  $\Delta 6$ ,  $\Delta 7$ ,  $\Delta 9p_{30}$ ,  $\Delta 9$ ,  $\Delta(9,10)$ ,  $\Delta 10$  y  $\Delta 11,12$ ), aunque sólo 6 son eventos confirmados ( $\Delta 7$ ,  $\Delta 9p_{30}$ ,  $\Delta 9$ ,  $\Delta 9,10$  y  $\Delta 11,12$ ). Con la excepción de  $\Delta 11,12$ ; todos son eventos que representan un % mínimo de la expresión de *PALB2*. Tanto los datos de RNAseq como un estudio (indirecto) mediante PCR digital sugieren que  $\Delta 11,12$  podría representar  $\approx 10\%$  de la expresión global. Los exones 11 y 12 (residuos 1039-1117) codifican parte del motivo WD4 y todo el motivo WD5, por lo que es muy probable que afecten adversamente al dominio C-terminal WD40 (residuos 867-1186) de *PALB2*, esencial para la interacción de *PALB2* con *BRCA2*, *RAD51D*, *RAD51C*, *XRCC3* y *POLn* (Buisson et al, 2014; Park et al, 2014). Por ello,

consideramos muy improbable que los transcritos  $\Delta 11,12$  codifiquen una proteína PALB2 funcional capaz de conferir haplosuficiencia. Consideraciones estructurales similares nos llevan a considerar que muy probablemente  $\Delta 2$ ,  $\Delta 2_5$ ,  $\Delta 4$ ,  $\Delta(4,5p_{24})$ ,  $\Delta 6$ ,  $\Delta 7$ ,  $\Delta 9p_{30}$ ,  $\Delta 9$ ,  $\Delta 9,10$  y  $\Delta 10$  no codifican proteínas funcionales. La pérdida de los exones 2 y 4 elimina regiones críticas para la unión a *BRCA1* y a DNA. La pérdida de los exones 6, 7, 9 y 10 (o la pérdida de los 30 primeros nucleótidos del exón 9) elimina alguna de las siete repeticiones WD en tándem que forman el dominio WD40 de *PALB2* (información consultada en UniProtKB - Q86YC2). En resumen, consideramos improbable que estos eventos de *splicing* alternativo tengan alguna relevancia en la clasificación clínica de variantes genéticas en *PALB2*. Por el contrario, pensamos que los eventos  $\Delta 2p_{6/9}$  y  $\Delta 5p_{24}$  pueden ser relevantes.

Las variantes localizadas en el aceptor del exón 2 de *PALB2* podrían producir transcritos r.49\_54del o r.49\_57del. Estos transcritos codifican proteínas que carecen de dos residuos p.(Leu17\_Lys18del), o de 3 residuos p.(Leu17\_Glu19) localizados en el dominio N-terminal (*coiled-coil domain*, residuos 9-41) que media la interacción de *PALB2* con *BRCA1* (Sy SM, et al., 2009; Zhang F et al., 2009). Los residuos 14, 21, 28, 35 y 42 son críticos para dicha interacción (Sy SM, et al., 2009), pero desconocemos el efecto preciso de la pérdida de los residuos 17, 18 y 19, y no podemos asumir que impliquen pérdida de función. Por ello, pensamos que las variantes en el aceptor de *splicing* del exón 2 de *PALB2* deben considerarse, en ausencia de estudios *in vitro* de *splicing*, Clase-3.

Del mismo modo, las variantes localizadas en el aceptor del exón 5 de *PALB2* podrían producir transcritos r.1685\_1708del que codifican para una proteína p.(Gly562\_Lys569del) que, hasta donde sabemos (información consultada en UniProtKB - Q86YC2), no afecta a ningún dominio y/o residuo crítico para la actividad de *PALB2*. Por ello, pensamos que, en ausencia de estudios *in vitro* de *splicing*, las variantes en el aceptor de *splicing* del exón 5 deben considerarse Clase-3.

Debido a su mínima contribución a la expresión global de *PALB2*, consideramos muy improbable que, aun codificando proteínas funcionales, los niveles normales de expresión de  $\Delta 2p_{6/9}$  o  $\Delta 5p_{24}$  confieran haplosuficiencia y puedan “rescatar” variantes PTC. Por ello, no vemos razones para no considerar Clase-5, todos los PTCs identificados en los exones 1 a 12 de *PALB2*. Recomendamos ser más conservadores con las variantes PTCs identificadas en el último exón (exón 13) del gen, aunque esto nada tiene que ver con el *splicing* alternativo.

Gen	Clase-3 <sup>1</sup>	Justificación <sup>2</sup>	Variantes Identificadas <sup>3</sup>	Patología <sup>4</sup>	Referencia <sup>5</sup>	Clasificación Publicada <sup>6</sup>
RAD51C	c.146-1,2 (IVS1-1,2)	$\Delta 2p_3$	-	-	-	-
	c.966-1,2 (IVS7-1,2)	$\nabla 8p^3$	c.966-1G>A	BRCA1	Ambry Genetics (ClinVar)	Clase-3 <sup>7</sup>
	c.1027-1,2 (IVS8-1,2)	$\Delta 9p_6$	c.1027-2A>G c.1027-2A>G	FANCO CO	Invitae (ClinVar) Desmond et al, 2015	Clase-4/5 Clase-4/5
RAD51D	c.83-1,2 (IVS1-1,2)	$\Delta 2p_{21}$	-	-	-	-
	PTC en E3 <sup>179</sup>	E3 <sup>179</sup>	-	-	-	-
	c.145-1,2 (IVS3-1,2)	E3 <sup>179</sup>	-	-	-	-
	c.263+1,2 (IVS3+1,2)	E3 <sup>179</sup>	c.263+1G>A	CO	Norquist et al, 2016	Clase-4/5
	PTC en E3 <sup>179</sup>	E3 <sup>179</sup>	c.161C>G (p.Ser54Ter) c.184C>T (p.Gln62Ter) c.216G>A (p.Trp72Ter) c.277dupT (p.Ser93PhefsTer7) c.293delA (p.Asp98ValfsTer25) c.293delA (p.Asp98ValfsTer25) c.320C>G (p.Ser107Ter)	Control Control Control Control Control S Ewing Control	gnomAD gnomAD gnomAD gnomAD gnomAD Brohl et al, 2017 gnomAD	- - - - - Clase-4/5 -
			-	-	-	-
			-	-	-	-
			-	-	-	-
			-	-	-	-
			-	-	-	-
	c.145-1,2	E3 <sup>179</sup>	-	-	-	-
	c.323+1,2	E3 <sup>179</sup>	-	-	-	-
PALB2	c.49-1,2 (IVS1-1,2)	$\Delta 2p_6$	c.49-1G>T c.49-2A>T	Control CM TN	gnomAD Couch F et al., 2015	- Clase-4/5
	c.1685-1,2 (IVS4-1,2)	$\Delta 5p_{24}$	c.1685-1G>C c.1685-2A>G c.1685-2A>G	- BRCA1 Control	Gene Dx (ClinVar) Li J et al, 2016 gnomAD	Clase-5 Clase-4/5 -

**Tabla 19. Implicaciones del *splicing* alternativo en la clasificación clínica inicial de variantes genéticas en**

**RAD51C, RAD51D y PALB2.** <sup>1</sup>Proponemos que las variantes localizadas en los donadores/aceptores de *splicing* indicados (en ausencia de estudios *in vitro* de *splicing* y/o estimaciones directas de riesgo) se deben tratar de forma conservadora (Clase-3). Para las variantes en otros donadores/aceptores apoyamos una clasificación inicial menos conservadora (Clase-4). <sup>2</sup>Se indica el evento de *splicing* alternativo que, en nuestra opinión, justifica ser conservador. <sup>3</sup>Se han buscado variantes genéticas en gnomAD, Flossies, ClinVar y artículos científicos accesibles a través de PubMed. Las variantes en RAD51D se describen respecto al transcrito de referencia ENST00000345365, salvo las variantes PTC en E3<sup>179</sup> que se describen respecto al transcrito ENST00000590016. <sup>4</sup>BRCA1 (Síndrome Mama/Ovario sin mutación en BRCA1/2), FANCO (Anemia de Fanconi, grupo de complementación O), CO (Cáncer de Ovario), S Ewing (Sarcoma de Ewing), CM TN (Cáncer de Mama Triple Negativo). <sup>5</sup>Ambry Genetics, Invitae y Gene Dx son laboratorios privados de diagnóstico genético que suben de forma regular su información a ClinVar. <sup>6</sup>Clasificación clínica propuesta en la publicación citada. Las variantes descritas en gnomAD no tienen clasificación clínica asociada <sup>7</sup>Aparece descrita por AmbryGenetics como Variante sin clasificar (Clase-3) en ClinVar, aunque los propios criterios de clasificación publicados por Ambry Genetics en ClinVar justificarían Clase-4 “Alterations at the canonical donor/acceptor sites(+/-1,2) without other strong evidence supporting pathogenicity”.

## 2.2. *Splicing* Alternativo y Diseño experimental de estudios *in vitro* de *splicing* en *RAD51C*, *RAD51D* y *PALB2*

En la sección anterior hemos discutido la relevancia de nuestros hallazgos (eventos de *splicing* alternativo) en relación a la clasificación clínica inicial de variantes genéticas en *RAD51C*, *RAD51D* y *PALB2*. Por clasificación clínica inicial nos referimos a la clasificación que se realiza en el momento de identificar la variante en un test genético (basándose exclusivamente en el tipo y localización de la variante), sin contar con datos adicionales tales como una estimación directa de riesgo y/o un estudio *in vitro* de *splicing*.

En el caso de variantes con alta probabilidad de afectar al *splicing*, el siguiente paso lógico es intentar realizar estudios *in vitro* (RT-PCR sobre RNA de portadores) o *ex vivo* (minigen) de *splicing*. Tal y como se comentó en la introducción (*BRCA1* c.591C>T), la selección de los cebadores es crítica para los estudios *in vitro* en portadores, y de hecho diversas combinaciones de cebadores pueden dar resultado completamente discrepantes (Dosil et al, 2010). Idealmente, la selección de cebadores debe ser tal que el estudio por RT-PCR permita capturar información de todas las isoformas de *splicing* alternativo potencialmente afectadas por la variante a interrogar. Por ello, el diseño de este tipo de estudios debe tener en cuenta la existencia de *splicing* alternativo. Del mismo modo, en los estudios *ex vivo* realizados en minigenes será crítico demostrar que la versión wt del minigen es capaz de reproducir adecuadamente el *splicing* alternativo.

De hecho, en los estudios por RT-PCR/EC de variantes en *BRCA1* y *BRCA2* presentados en esta tesis doctoral hemos tenido muy presente los eventos de *splicing* alternativo descritos en estos genes (Colombo et al, 2014; Fackenthal et al, 2016). Por ejemplo, la variante *BRCA1* c.594-2A>C (IVS9-2A>C) se analizó con cebadores en los exones 8 y 11 para poder detectar *BRCA1* Δ9,10, *BRCA1* c.4675+24G>A (IVS15+24G>A) se analizó con cebadores en los exones 14 y 19 para poder detectar *BRCA1* Δ15\_18, *BRCA1* c.5333-2A>G (IVS21-2A>G) se analizó con cebadores en los exones 20 y 24 para poder detectar *BRCA2* Δ22,23, y *BRCA2* c.7007G>A (p.Arg2336His) se analizó con cebadores en los exones 11 y 14 para poder detectar *BRCA2* Δ12,13.

Tal y como se indica en la tabla 20, se han identificado numerosos eventos de *splicing* alternativo multi-cassette en *RAD51C*, *RAD51D* y *PALB2* que pensamos deberían tenerse en cuenta al diseñar estudios de *splicing* en estos genes. Particularmente significativos nos parecen los eventos que implican mayor número de exones consecutivos, como *RAD51C* Δ6\_8, *RAD51D* Δ2\_5, Δ3\_6, y *PALB2* Δ2\_5. En nuestra opinión, la existencia de estos eventos nos

puede obligar a posicionar los cebadores de un estudio RT-PCR muy alejados del nucleótido de interés (o a diseñar minigenes que incluyan un número elevado de exones). Por ejemplo, pensamos que los estudios *in vitro* de *splicing* de una variante en el aceptor de *splicing* del exón 6 de *RAD51C* debe realizarse con cebadores que se localicen al menos en los exones 5 y 9.

Se han publicado numerosos estudios *in vitro/ex vivo* de *splicing* en los genes *RAD51C*, *RAD51D* y *PALB2*. En la tabla 20 resumimos algunos de estos estudios y nos preguntamos si el *splicing* alternativo descrito en nuestro trabajo podría matizar alguno de los resultados. La tabla también apoya parcialmente una de las hipótesis de trabajo de esta tesis doctoral: con frecuencia, los alelos portadores de mutaciones de *splicing* no expresan nuevos transcritos, si no que la incapacidad de expresar el transcrito de referencia “desplaza la expresión mayoritaria” a otras isoformas de *splicing* alternativo ya expresadas de forma minoritaria a partir de alelos wt.

Gen	Variante	D/A <sup>1</sup>	Efecto <sup>2</sup>	RT-PCR <sup>3</sup>	Referencia	<i>Splicing</i> Alternativo No Testado <sup>4</sup>
<b>RAD51C</b>	c.145+1G>T	D-E1	$\Delta 1q_{156} + \Delta 1q_{103}$ (SA)	E1-E3	Meindl et al, 2010	-
	c.404G>C	D-E2	$\nabla 2q^{27}$ (SA)	E1-E3	Neidhardt et al, 2016	-
	c.404G>T	D-E2	$\nabla 2q^{27}$ (SA)	E1-E3	Neidhardt et al, 2016	-
	c.706-2A>G	A-E5	$\Delta 5$ (SA)	E3-E7	Goldmar et al, 2013	-
	c.904+5G>T	D-E6	$\Delta 6$ (SA)	E5-E7	Meindl et al, 2010	$\Delta 6,7 + \Delta 6\_8$
	c.905-2A>G	A-E7	$\Delta 7$ (SA)	E7 (Minigen)	Coulet et al, 2013	$\Delta 6,7 + \Delta 6\_8 + \Delta 7,8$
	c.965+5G>A	D-E7	$\Delta 7$ (SA)	E7 (Minigen)	Coulet et al, 2013	$\Delta 6,7 + \Delta 6\_8 + \Delta 7,8$
<b>RAD51D</b>	c.1026+5_1026+7del	D-E8	$\Delta 8$ (SA)	E5-E9 E5-E9	Goldmar et al, 2013 Janatova et al, 2015	$\Delta 8,9 + \nabla 10$
	c.577-2A>G	A-E7	$\nabla 7p^{41}$	?	Kraus et al, 2017	$\Delta 7,8?$
<b>PALB2</b>	c.667+2_667+23del	D-E7	$\Delta 7$ (SA)	E5-E10	Gutierrez-Enriquez et al, 2014	-
	c.48G>A	D-E1	$\Delta 1q_{17}$ (SA)	E1-E2	Cattuci et al, 2014	-
	c.109-12T>A	A-E3	$\nabla 3p^{10}$	?	Kraus et al, 2017	$\Delta 2\_5?$
	c.2559C>T	<i>de novo</i>	$\Delta 6p_{29}$	?	Casadei et al, 2011	No Aplica#
	c.3113G>A	D-E10	$\Delta 10 + \Delta 10p_{31}$	?	Casadei et al, 2011	$\Delta 9,10?$
	c.3113+5G>C	D-E10	$\Delta 9,10$ (SA)	E8-E12	Reid et al, 2007	-
	c.3350+4A>G	D-E12	$\nabla 12q^4$	E11-E12?	Reid et al, 2007	$\Delta 11,12$

**Tabla 20. Implicaciones del *splicing* alternativo en el diseño de estudios *in vitro/ex vivo* de *splicing* en *RAD51C*, *RAD51D* y *PALB2*.** La tabla resume algunos de los estudios *in vitro* de *splicing* realizados en portadores de variantes en estos genes (identificados a través de PubMed). <sup>1</sup>Se indica el donador (D) o aceptor (A) afectado por la variante. proponemos que las variantes indicadas (en ausencia de estudios *in vitro* de *splicing* y/o estimaciones directas de riesgo) se deben tratar de forma conservadora (Clase-3). <sup>2</sup>Se indica el efecto descrito de acuerdo a la referencia indicada. Con la excepción de *RAD51D*  $\nabla 7p^{41}$ , *PALB2*  $\nabla 3p^{10}$ , *PALB2*  $\Delta 10p_{31}$  y *PALB2*  $\nabla 12q^4$ , todos los efectos se pueden describir como aumento de los niveles de expresión de isoformas de *splicing* alternativo (SA) existentes en controles (la mayoría no detectables en agarosa). <sup>3</sup>Combinación de cebadores utilizada en la RT-PCR (o los exones presentes en el minigene). Con el signo de interrogación señalamos trabajos de los que no ha sido imposible extraer dicha información. <sup>4</sup>Se indican los eventos de *splicing* alternativo que, en nuestra opinión, deberían haber sido

testados para asegurar la calidad y la interpretación clínica del estudio *in vitro/ex vivo*. #En el caso de variantes que crean un sitio consenso de *splicing de novo*, el *splicing* alternativo en controles deja de tener relevancia.

En la práctica clínica, la inmensa mayoría de estudios *in vitro* de *splicing* se realizan a partir de muestras derivadas de sangre. Por supuesto, todo lo comentado hasta ahora en la discusión, así como la información contenida en las tablas 19 y 20, no tienen ningún sentido si el *splicing* alternativo de *RAD51C*, *RAD51D* y/o *PALB2* tuviese un fuerte componente tejido específico. En el presente trabajo hemos realizado una primera aproximación cualitativa a esta cuestión, combinando datos de RNAseq realizados en tejido normal de mama y fimbria ovárica con experimentos de RT-PCR/EC realizados en tejido normal de mama y ovario (los datos se resumen en las tablas 12, 15, 18 de resultados). Es importante señalar que estos estudios presentan numerosas limitaciones técnicas, entre las que destacan: 1) la baja cobertura de los datos de RNAseq obtenidos en mama y muy especialmente en fimbria ovárica, 2) los datos de RT-PCR/EC se refieren a dos únicas muestras de RNA comercial de mama y ovario, y 3) los datos de RNAseq y RT-PCR/EC contienen poca o ninguna información cuantitativa.

Por ello, pensamos que el hecho de que hayamos detectado menos eventos de *splicing* alternativo en mama/ovario (especialmente en ovario) que en LCLs/sangre no es una evidencia de *splicing* alternativo tejido específico, tan sólo un reflejo de la baja cobertura de nuestros análisis en mama/ovario. Sin embargo, pensamos que nuestros datos si tienen calidad suficiente para descartar la existencia de eventos de *splicing* alternativo mayoritarios específicos de mama/ovario. De algún modo, este es el dato fundamental que nos permite validar LCLs/sangre como fuente de RNA adecuada para realizar estudios *in vitro* de *splicing* en estos genes.

En general, nuestros datos no permiten detectar efectos tejido específicos cuantitativos (detección de los mismos eventos con distintas contribuciones relativas en distintos tejidos). Sin embargo, datos obtenidos por PCR digital aportan alguna información. En concreto, sugieren que la contribución relativa de los exones E3<sup>179</sup> y E3<sup>119</sup> a la expresión de *RAD51D* podría ser distinta en distintos tejidos (Figura 29), y la contribución relativa de *RAD51D* Δ3\_5 podría ser particularmente elevada en ovario (Figura 38). Sería necesario realizar este tipo de análisis en un número relevante de muestras independientes para poder confirmarlo.

En resumen, en nuestra opinión la existencia de *splicing* alternativo, de forma un tanto sorprendente, tiende a ser ignorada tanto en los estudios de investigación en susceptibilidad genética al cáncer de mama y ovario como en su traslación a la clínica. La presente tesis

doctoral defiende que esto es un error (en particular, cuando los resultados de la investigación se trasladan a la clínica), y postula que una caracterización detallada del *splicing* alternativo de un gen aporta información esencial para la correcta interpretación clínica de las variantes genéticas identificadas en dicho gen.

# CONCLUSIONES





# CONCLUSIONES

---

1. La incorporación a la rutina diagnóstica de los estudios *in vitro* de *splicing* permite reducir de forma significativa la incertidumbre asociada a la clasificación clínica de variantes genéticas en *BRCA1* y *BRCA2* (en particular, en variantes sinónimas e intrónicas).
2. Los estudios realizados sobre la variante *BRCA1* c.594-2A>C (IVS9-2A>C) nos permiten concluir que:
  - Los niveles normales de expresión de *BRCA1*  $\Delta 9,10$  confieren haplosuficiencia.
  - En los exones 9 y 10 de *BRCA1*, una variante en sitio consenso de *splicing* y/o un codón de parada prematuro no es necesariamente patogénico y, en ausencia de datos complementarios, deberá considerarse de significado clínico incierto.
3. Las variantes genéticas que causan pérdida completa del exón 3 de *BRCA2* son patogénicas. No obstante, los estudios realizados en la variante *BRCA2* c.68-7T>A (IVS2-7T>A) nos permiten concluir que niveles de  $\Delta 3$  hasta cuatro veces superiores a los observados en controles no se asocian con aumento del riesgo de desarrollar cáncer.
4. La caracterización del *splicing* alternativo en los genes *RAD51C*, *RAD51D* y *PALB2* nos permite proponer que:
  - En ausencia de datos complementarios, las variantes genéticas localizadas en los donadores consenso de *splicing* de los exones 2, 8 y 9 de *RAD51C*, 2 de *RAD51D*, y 2 y 5 de *PALB2* se consideren de significado clínico incierto.
  - En cualquiera de los dos exones 3 alternativos de *RAD51D*, una variante en sitio consenso de *splicing* y/o un codón de parada prematuro no es necesariamente patogénico y, en ausencia de datos complementarios, se consideren de significado clínico incierto
5. En conjunto, nuestros estudios demuestran que la caracterización detallada del *splicing* alternativo de un gen aporta información esencial para la correcta interpretación clínica de las variantes genéticas que se identifican en dicho gen, y es clave para el diseño de los estudios de *splicing*.



Anexos



# ANEXO I

## 1. Selección de Transcritos de Referencia para la anotación de eventos de *splicing* alternativo

### **RAD51C**

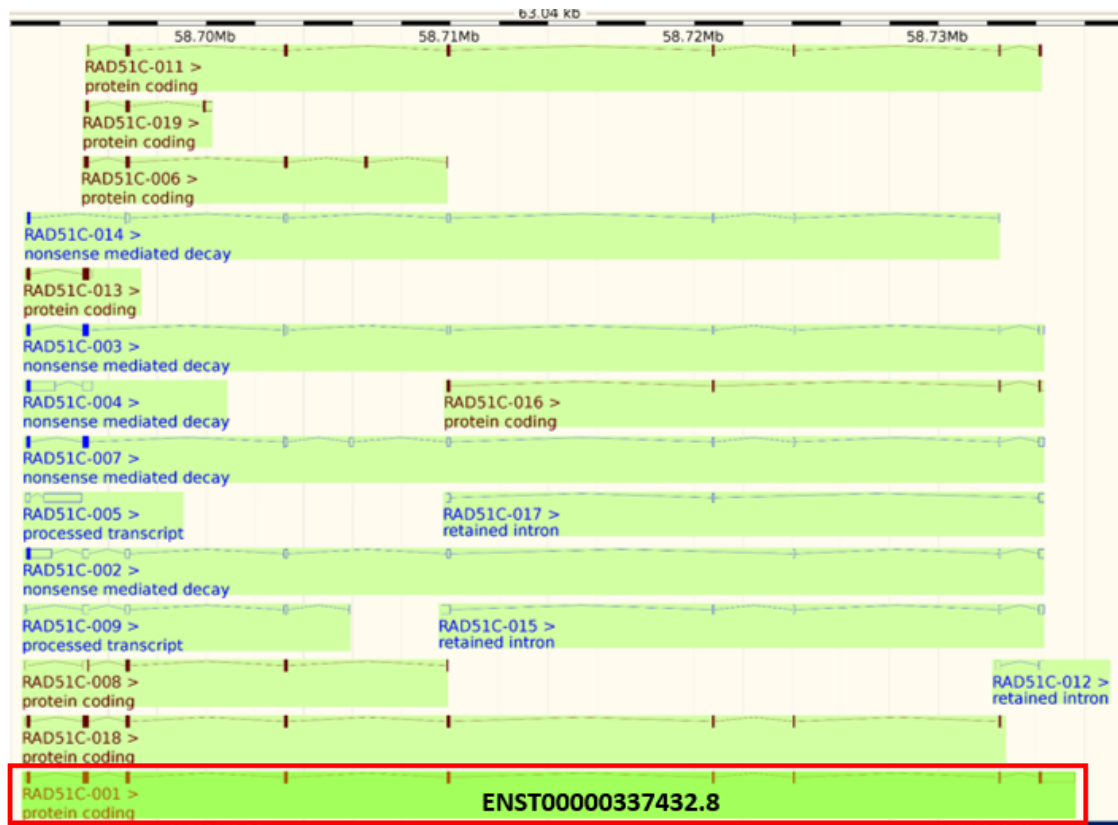


Figura 51. Transcritos del locus *RAD51C* descritos en ENSEMBL. En rojo se resalta el transcrito de referencia seleccionado.

**RAD51D**

En el caso del gen *RAD51D*, la selección del transcrito de referencia no es evidente, ya que existen dos transcritos prácticamente idénticos que sólo difieren en el uso de dos exones 3 alternativos de 119 ( $E3_{119}$ ) y 179 nucleótidos ( $E3_{179}$ ).

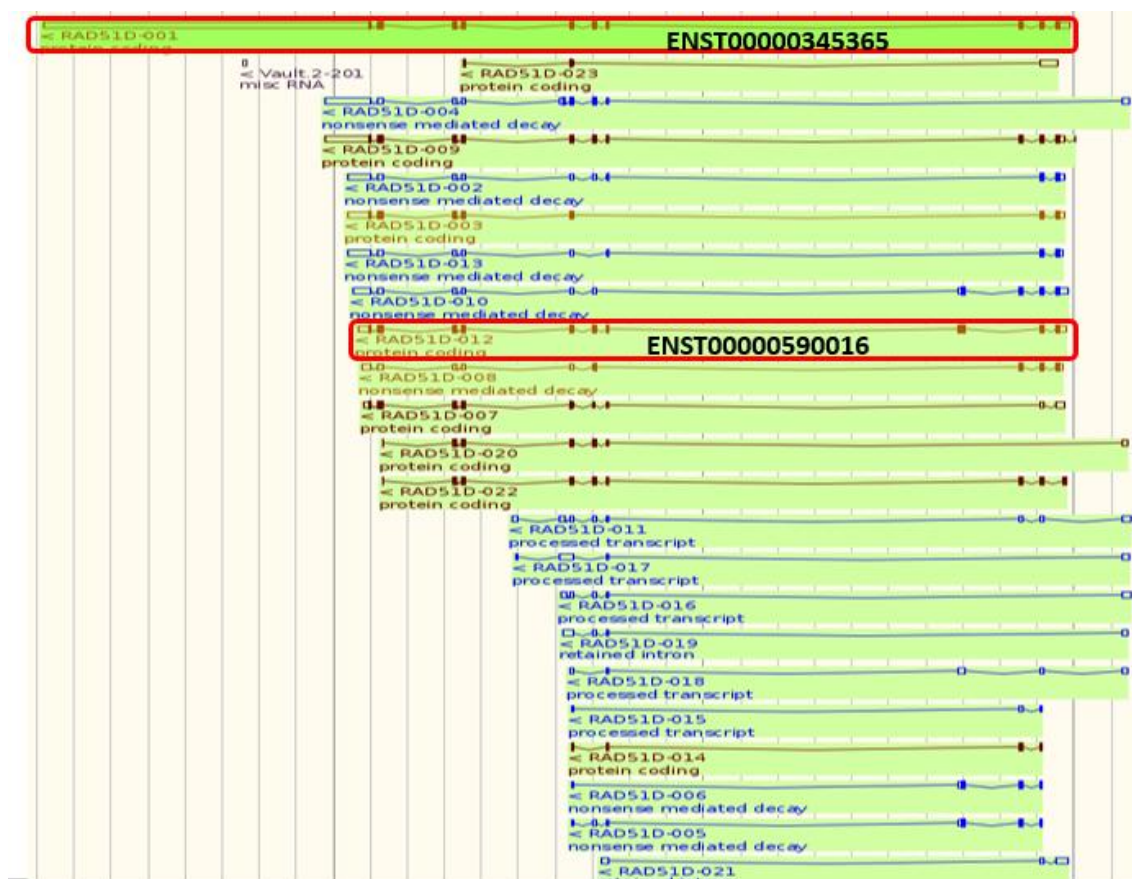


Figura 52. Transcritos del locus *RAD51D* descritos en ENSEMBL. En rojo se resaltan los transcritos de referencia empleados en la caracterización.

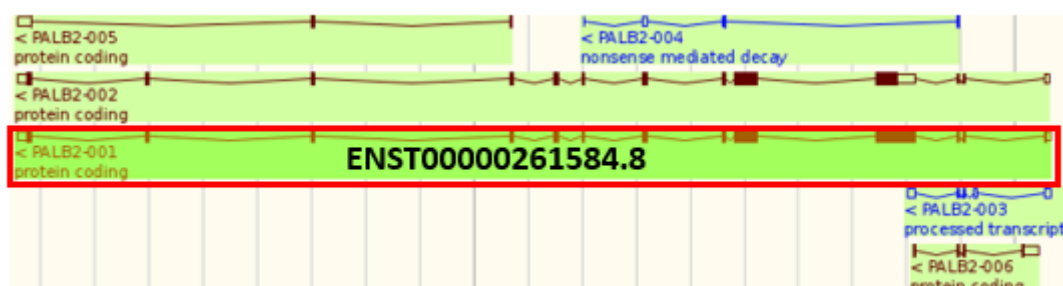
**PALB2**

Figura 53. Transcritos del locus *PALB2* descritos en ENSEMBL. En rojo se resalta el transcrito de referencia empleado en la caracterización.

### 1.1. Cebadores en el locus *BRCA1*

Los cebadores empleados para la caracterización de variantes localizadas en el locus *BRCA1* se han diseñado utilizando como transcrito de referencia ENSEMBL [ENST00000357654.7](https://www.ensembl.org/ENST00000357654.7).

Cebador	Región cDNA (sentido 5' a 3')	Cebador	Región cDNA (sentido 5' a 3')
1s	GACAGGCTGTGGGTTTCT	6a	CAGCTTGACACAGGTTTGGA
2s	ATGGATTATCTGCTCTTCGC	7a	TGGGCTACAGAAACCGTGCCA
3s	TCAAGGAACCTGTCTCCACA	8a	ACAAAGCAGCGGATACAACC
8s	TGTCCAACCTCTAACCTTGAACTGTG	11a	CAAACCTGCATGTGGAGCCA
10s	CCAGGGATGAAATCAGTTTGGA	14a	CCAGAATCCAGAAGGCCTTT
11s	TGTGAATTTCTGAGACGGATG	15int a	TAAGCCCACTTTAGAAAAATAGCTC
13s	ATGGGAGCCAGCCTTCTAAC	16a	GCAATGGAAGAAAGTGTGAG
14s	TCTGCAGATAGTTCTACCAG	16.1a*	GGGGAACTTTCAATGCAGAG
16s	AAAGAATGTCCATGGTGGTG	17.1a*	TTTATGCAGCAGATGCAAGG
16.1s	TGAAGACAGAGCCCCAGAGT	19a	AAAGAAAGAAAAATGCTGAATGA
20s	GGAGATGTGGTCAAATGGA	22a	AATGGATGGTACAGCTGTG
		24a	GAGTGGGTGTTGGACAGTGTAGCA

**Tabla 21.** Cebadores utilizados en los estudios de *splicing* por electroforesis capilar y secuenciación Sanger del locus *BRCA1*. Los cebadores *antisentido* (en inglés, *antisense*) están marcados con un fluoróforo (FAM).

\*Cebadores *antisense* que no han sido marcados con FAM para facilitar la secuenciación Sanger



## 1.2. Cebadores en el locus *BRCA2*

El diseño de los cebadores para la caracterización de variantes localizadas en el locus *BRCA2* se ha empleado el transcrito de referencia ENSEMBL: [ENST00000380152.7](#).

Cebador	Región cDNA (sentido 5' a 3')	Cebador	Región cDNA (sentido 5' a 3')
1s	AGCTTACTCCGGCCAAAAA	4a	GGACAGGAAACATCATCTGCT
2s	GGATCCAAAGAGAGGCCAAC	6a	TGAAACAACTCCCACATACCA
3s	CTCCACAAAGGAAACCATCTT	8a	ATCATGAGGAAATACAGTTTCAGAT
4s	AGCAGATGATGTTTCCTGTCC	10a	TGACTGTGGTGAAACGAG
7s	GGTCGTCAGACACCAAAACA	10.1a*	GCGTTTGCTTCATGGAAAAAT
9s	TGTGACAGACAGTGAAAACACAA	10.2a*	TTCCACTCTCAAAGGGGCTTC
10s	CTAATTACTGTTTCAGCCAG	11 <sup>a</sup>	CACAATTGTCCCAAAAGAGC
10.5s	TGAAGAGCAGCATCTTGAATCT	12 <sup>a</sup>	GCCATCTGGAGTGCTTTTGTG
11s	CCAAGTCATGCCACACATTC	14 <sup>a</sup>	AGTTGTGCAAAGGGTACACA
11 <sub>25-26</sub> s	GCCATTGAGACCAGCTCAAA	14.1a*	GTTCAACTCTGTGAAAATGAG
11.29s	GCAGGCTTCACCTAAAAACG	15.1a*	GCTGTGGAAAGACGCAAAGCC
12s	AGGCTTCAAAAAGCACTCCA	15.2a*	TTCGCATATCCTGTATATCTCTGG
14.2s	GAGACACTTGATTACTACAG	17a	CCATAGCTGCCAGTTCCAT
15.2s	GAATGCCAGAGATATACAG	17.1a*	ACAAAACTCATTGTTCAAGATCAT
18s	GAAGCCCCAGAATCTCTTATG	19a	CAAGTTTGGTATACCAGCGAG
19s	CATCGCTTTTCAGTGATGGA	20a	TTGTTGGGCTCCACATATT
20s	ACTAAAATTCAGGAGGAATTGAAG	22a	CTCTTAAGTCTCTTCACTG
22s	TAGGAAGGCCATGGAATCTG	23a	CTTTTAGATTTTGAAGTTGCAAGATG
		24a	CCGTGGCTGGTAAATCTGA
		25.1a*	GCCTGATTGGATTCTGGTC

**Tabla 22.** Cebadores utilizados en los estudios de *splicing* por electroforesis capilar y secuenciación Sanger del locus *BRCA2*. Los cebadores *antisense* están marcados con un fluoróforo (FAM). \*Cebadores *antisense* que no han sido marcados con FAM para facilitar la secuenciación Sanger.

## 1.3. Cebadores en el locus *RAD51C*

Cebador	Región cDNA (sentido 5' a 3')	Cebador	Región cDNA (sentido 5' a 3')
1s	GAGGGCGTGCGGAGTTTG	2a	TCTAGTGCTGAACAGAAGGTGA
1s-interno	AAACCTCCGAGCTTAGCAA	2a-interno	TGCAAGAGTTTCTAAGGCTTCTG
2s	GAGCATACCCAGGGCTTCAT	3a	CTGCCACTCTCCAAA
3s	AGATGTGCAGATACCAGAATGT	4a	AGTGTCTGAAAGGAAATCTGGA
4s	GCTTTGGAGGATTTCACTCTTGA	5a	GATCATTTGCTGGGCTAGGC
5s	CCTAGATGACCTGTCTCTTCGT	7a	GCCGRARRGTAGCAGCATGT
7s	GGGACATGCTGTACAATACG	8a	TGGGTGACTTGTACAAT
		9a	TTCTTCTCTGGGTCTCGTG

**Tabla 23.** Cebadores utilizados en los estudios de *splicing* por electroforesis capilar y secuenciación Sanger del locus *RAD51C*. Los cebadores *antisense* están marcados con un fluoróforo (FAM).

#### 1.4. Cebadores en el locus *RAD51D*

Cebador	Región cDNA (sentido 5' a 3')	Cebador	Región cDNA (sentido 5' a 3')
1s	GACTGTGCCCTGGCCTTAC	4a	CTGGGCCTCCTACAATTTCA
2s	GACCTGGTTTCTGCAGACCT	5a	CATCCTGGGTTTTAGCCTGA
3s (corto)	GAATGGCGCTGATCTCTACG	7a	TCTGCTGACCTCCAGAAGT
3s (largo)	CTGAAGAAGGCAGGACTTGG	8a	GCCAGGCTCTTCAGCTCTC
4s	AAATTGTAGGAGGCCAGGT	9a	TCGATGGTGTCCAGGAGAAT
6s	GGTGCATGCATTTGACATCT	10a	AGTGCCAGGTGGCAGTAAAC

**Tabla 24.** Cebadores utilizados en los estudios de *splicing* por electroforesis capilar y secuenciación Sanger del locus *RAD51D*. Los cebadores *antisense* están marcados con un fluoróforo (FAM).

#### 1.5. Cebadores en el locus *PALB2*

Cebador	Región cDNA (sentido 5' a 3')	Cebador	Región cDNA (sentido 5' a 3')
1s	GTCAGCTGATCGCGCACT	3a	GAGTGTTTTAGCTGCGGTGA
3s	CGTGCCCAAAGAGCTGAA	5a	AAAGGGCTCCACTGGTTTTT
4s	GCACAGGACAACCAAGTTCA	6a	GGTTGCCTGGGTTTATGCTA
5s	GGACAACCTACCTGTGACTGTG	8a	TTCCCAAAGCTACACACACG
6s	AGCAGAGCTTCCTGCTTCTG	10a	AAGCAGAGCTTCTTGATCC
7s	CCGGTTGTAAAGAGCCATGT	11a	CTTTGTGACAGACTGAAGC
9s	GGCATTGTTTTGTTCTCTGA	12a	TGCCCTGGAGGAAGACAGTA
10s	ATTTTGTATGCCCCCTGAG		

**Tabla 25.** Cebadores utilizados en los estudios de *splicing* por electroforesis capilar y secuenciación Sanger del locus *PALB2*. Los cebadores *antisense* están marcados con un fluoróforo (FAM).

## 2. Análisis Cuantitativo del *splicing* alternativo mediante PCR digital

### Cuantificación de $\Delta 9,10$ respecto a la expresión global de *BRCA1*

El ensayo marcado con FAM reconoce la unión exón 8 y exón 11 (E8E11), y por tanto detecta el evento de *splicing* alternativo  $\Delta 9, 10$ . El ensayo marcado con VIC reconoce la unión exón 23 y exón 24 (E23E24). Dado que no se han descrito isoformas de *splicing* alternativo en esta región del gen, asumimos que la sonda VIC mide la expresión global del gen *BRCA1* (Figura 54).



**Figura 54.** Diseño de los ensayos TaqMan para cuantificar la isoforma  $\Delta 9, 10$  de *BRCA1*. El ensayo TaqMan marcado con VIC (en rojo) reconoce la unión exón23-exón24, que consideramos un correlato aproximado de la expresión global del locus *BRCA1* (con la excepción de BRCA-IRIS); mientras que el ensayo TaqMan marcado con FAM (en azul) representa la isoforma de *splicing*  $\Delta 9,10$  de *BRCA1*.

### Cuantificación de la unión E9E10 respecto a la expresión global de *BRCA1*

El ensayo TaqMan marcado con FAM reconoce la unión exón9-exón10 se combina con el ensayo VIC que reconoce la unión 23-24.



**Figura 55.** Diseño del ensayo TaqMan para cuantificar la expresión de los exones 9 y 10 de *BRCA1*. El ensayo TaqMan marcado con VIC (en rojo) reconoce la unión exón23-exón24, que consideramos un correlato aproximado de la expresión global del locus *BRCA1* (con la excepción de BRCA-IRIS); mientras que el ensayo TaqMan marcado con FAM (en azul) reconoce la unión exón 9-10.

### Cuantificación de $\Delta 3$ respecto a la expresión global de *BRCA2*

Para la caracterización cuantitativa de este evento de *splicing* se han diseñado dos ensayos Taqman (Figura 56). El ensayo TaqMan marcado con FAM representa la expresión del evento  $\Delta 3$  y ha sido diseñado colocando un cebador *sense* en el exón 2 y un cebador *antisense* en el exón 4. Como referencia se ha diseñado un ensayo TaqMan marcado con VIC que representa la unión de los exones 26 y 27. Este ensayo ha sido diseñado, colocando un cebador *sense* en el exón 26 y un cebador *antisense* en el exón 27.



**Figura 56.** Diseño de los ensayos TaqMan para cuantificar la expresión del evento  $\Delta 3$  con respecto a la expresión global del locus *BRCA2*. El ensayo TaqMan marcado con FAM (en azul) reconoce la unión de los exones 2 y 4, mientras que el marcado con VIC (en rojo) reconoce la unión de los exones 26 y 27, que consideramos un correlato aproximado de la expresión global del locus *BRCA2*.

### Cuantificación de $\Delta 3$ respecto a la unión E3E4 de *BRCA2*

Además, la isoforma  $\Delta 3$  de *BRCA2* también ha sido cuantificada con respecto a la unión de los exones 3 y 4. Para ello se empleó la sonda marcada con FAM que hibrida en los exones 2 y 4, y, por lo tanto, representará el *skipping* del exón 3 ( $\Delta 3$ ). Y la sonda de referencia marcada con VIC, hibrida en los exones 3 y 4 para representar la unión de ambos exones (Figura 57).

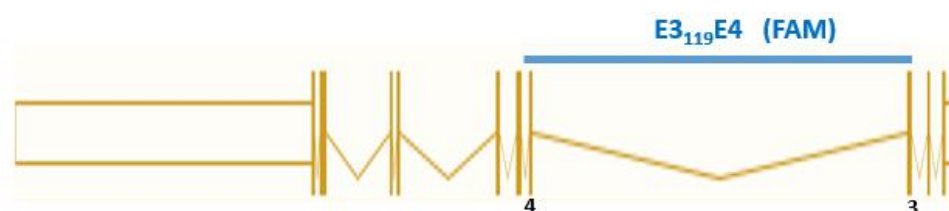


**Figura 57.** Diseño de los ensayos TaqMan para cuantificar la expresión del evento  $\Delta 3$  con respecto a la unión exón 3 y exón 4 del locus *BRCA2*. El ensayo TaqMan marcado con FAM (en azul) representa la unión de los exones 2 y 4 de *BRCA2*; mientras que el marcado con VIC (en rojo) reconoce la unión de los exones 3 y 4 de *BRCA2*.

### Cuantificación de los dos exones 3 alternativos de *RAD51D*

Para el exón 3 de 119 nucleótidos ( $E3_{119}$ ) se ha diseñado un ensayo TaqMan marcado con FAM que representa la señal del exón 3<sub>119</sub> (reconoce la unión del exón 3<sub>119</sub> con el exón 4, Figura 58). Y para el exón 3 de 179 nucleótidos ( $E3_{179}$ ), se ha diseñado un ensayo TaqMan marcado con FAM (reconoce la unión del exón 3<sub>179</sub> con el exón 4, Figura 59).

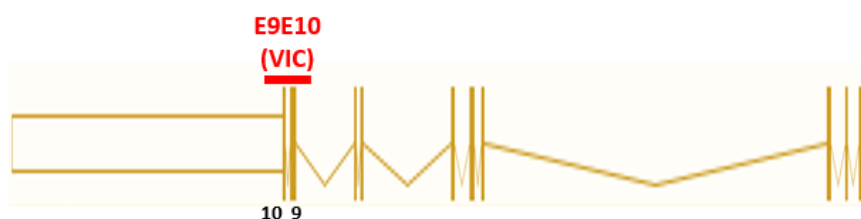
Como sonda TaqMan de referencia (marcada con VIC), se ha diseñado un ensayo que reconozca la unión de los exones 9-10 (Figura 60).



**Figura 58.** Diseño de los ensayos TaqMan para cuantificar la expresión del exón 3<sub>119</sub> del locus *RAD51D*. El ensayo TaqMan marcado con FAM (en azul) representa la unión de los exones 3<sub>119</sub> y 4 de *RAD51D*.



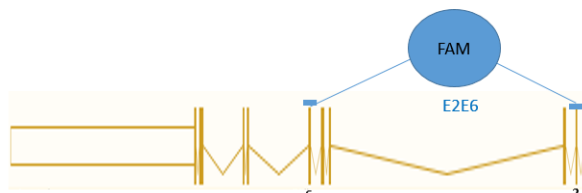
**Figura 59.** Diseño de los ensayos TaqMan para cuantificar la expresión del exón 3<sub>179</sub> del locus *RAD51D*. El ensayo TaqMan marcado con FAM (en azul) representa la unión de los exones 3<sub>179</sub> y 4 de *RAD51D*.



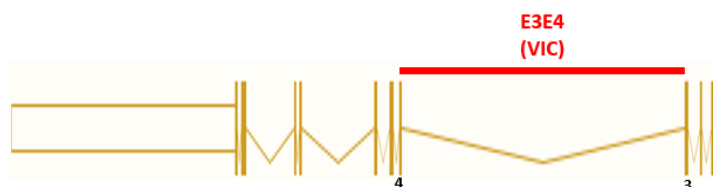
**Figura 60.** Diseño de los ensayos TaqMan para cuantificar la expresión de la unión de los exones 9 y 10 del locus *RAD51D*. El ensayo TaqMan marcado con VIC (en rojo) representa la unión de los exones 9-10 de *RAD51D*.

### Ensayo TaqMan para cuantificar el evento de *RAD51D* $\Delta 3\_5$

Para cuantificar la contribución de  $\Delta 3\_5$  a la expresión de *RAD51D*, se han diseñado dos ensayos TaqMan. Uno marcado con FAM, que representa la expresión del evento  $\Delta 3\_5$  (Figura 61). Y otro ensayo marcado con VIC, que reconoce la unión de los exones 3-4 (Figura 62).



**Figura 61.** Diseño de los ensayos TaqMan para cuantificar la expresión del evento  $\Delta 3\_5$  del locus *RAD51D*. El ensayo TaqMan marcado con FAM (en azul) captura la unión de los exones 2 y 6 de *RAD51D* ( $\Delta 3\_5$ ). El primer *sense* hibrida en el exón 2 y el primer *antisense* en el exón 6 de *RAD51D*.



**Figura 62.** Diseño de los ensayos TaqMan para cuantificar la expresión de la unión de los exones 3 y 4 del locus *RAD51D*. El ensayo TaqMan marcado con VIC (en rojo) captura la unión de los exones 3-4 de *RAD51D*. Se ha colocado un cebador *sense* en el exón 3 y un *antisense* en el exón 4 de *RAD51D*.

### Cuantificación de *PALB2* $\Delta 4,5$ vs. E4E5

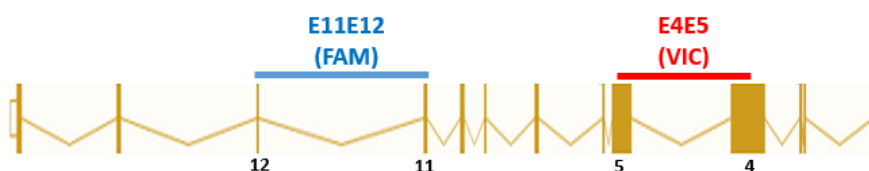
Para cuantificar la contribución del evento  $\Delta 4,5$  se han diseñado dos ensayos TaqMan (Figura 63). El ensayo marcado con FAM reconoce la unión de los exones 3 y 6, por lo que representa el evento de *splicing*  $\Delta 4,5$ . Mientras que el ensayo marcado con VIC reconoce la unión de los exones 4 y 5, que es el ensayo empleado como referencia.



**Figura 63.** Diseño de los ensayos TaqMan para cuantificar la expresión del evento  $\Delta 4,5$  respecto a la expresión de la unión de los exones 4 y 5 del locus *PALB2*. El ensayo TaqMan marcado con FAM (en azul) representa la unión de los exones 3 y 6 de *PALB2*; mientras que el marcado con VIC (en rojo) reconoce la unión de los exones 4 y 5 de *PALB2*.

### Cuantificación de *PALB2* $\Delta 11,12$

Se ha cuantificado de manera indirecta la contribución del evento  $\Delta 11,12$ . Para ello se han diseñado dos ensayos TaqMan (Figura 64). El ensayo marcado con FAM reconoce la unión de los exones 11 y 12. Mientras que el ensayo marcado con VIC reconoce la unión de los exones 4 y 5, que es el ensayo empleado como referencia.



**Figura 64.** Diseño de los ensayos TaqMan para cuantificar la expresión del evento  $\Delta 11,12$  respecto a la expresión de la unión de los exones 4 y 5 del locus *PALB2*. El ensayo TaqMan marcado con FAM (en azul) representa la unión de los exones 11 y 12 de *PALB2*; mientras que el marcado con VIC (en rojo) reconoce la unión de los exones 4 y 5 de *PALB2*.

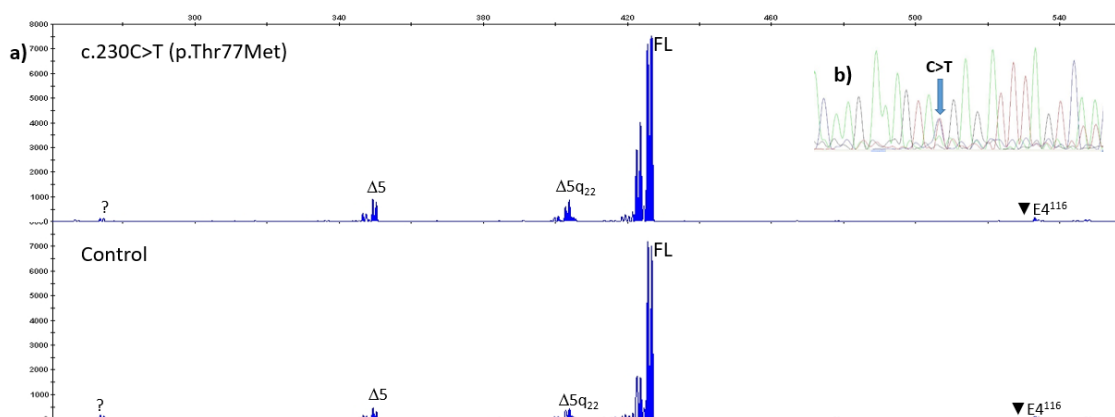
## ANEXO II

### 1. Análisis de *splicing* en portadores de variantes genéticas en *BRCA1*

#### **BRCA1 c.230C>T (p.Thr77Met)**

Variante *missense* localizada en el exón 6 de *BRCA1*. Hasta donde sabemos, no se han descrito estudios de *splicing* previos.

Para caracterizar el posible efecto de la variante a nivel de *splicing* se realizó un estudio RT-PCR/EC con cebadores localizados en los exones 3 y 8, a partir de RNA de sangre periférica obtenido de un portador y controles sanos (Figura 65a). No se observaron diferencias cualitativas. Tanto en controles como en el portador de la variante se detectó un producto predominante correspondiente al transcrito de referencia, y varios productos minoritarios, incluyendo  $\Delta 5$ ,  $\Delta 5q$  y  $\nabla 4^{116}$  (todos ellos descritos previamente en controles, Colombo et al, 2014). Todos los productos se detectan como un doblete de dos picos diferenciados en 3 nucleótidos, debido a la presencia de  $\Delta 8p_3$  (evento descrito previamente en controles, Colombo et al, 2014). La secuenciación de los productos de la RT-PCR (Figura 65b) demostró expresión bialélica balanceada en el portador de la variante.



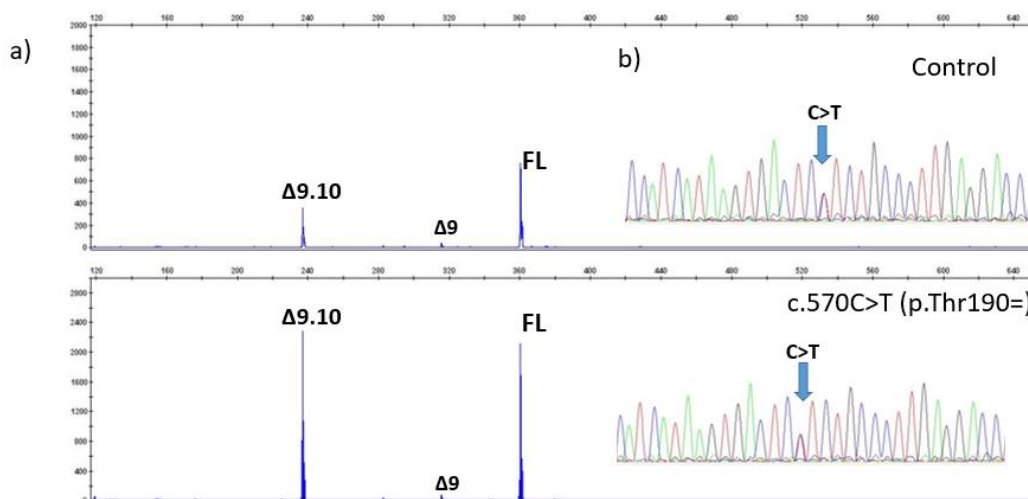
**Figura 65. Análisis del *splicing* de la variante *BRCA1* c.230C>T (p.Thr77Met).** a) Análisis del *splicing* mediante RT/PCR-EC. Electroferograma en el que se observa la contribución relativa de los eventos de *splicing*  $\Delta 5$  (404nt),  $\Delta 5q_{22}$  (402nt), transcrito de referencia (FL, 424nt) y  $\nabla E4^{116}$  a la expresión global detectada con el ensayo RT-PCR E3E8, medido en controles sanos de sangre, y en el portador de la variante c.230C>T (p.Thr77Met). b) Secuenciación por Sanger para estudiar expresión bialélica. El producto analizado en fragmentos se secuenció con un cebador *sense* E3.



**BRCA1 c.528G>A (p.Thr176=), c.570C>T (p.Thr190=), c.591C>T (p.Cys197=), c.773C>G (p.Pro258Arg), c.946A>G (p.Ser316Gly) y c.1242C>T (p.Asp414=)**

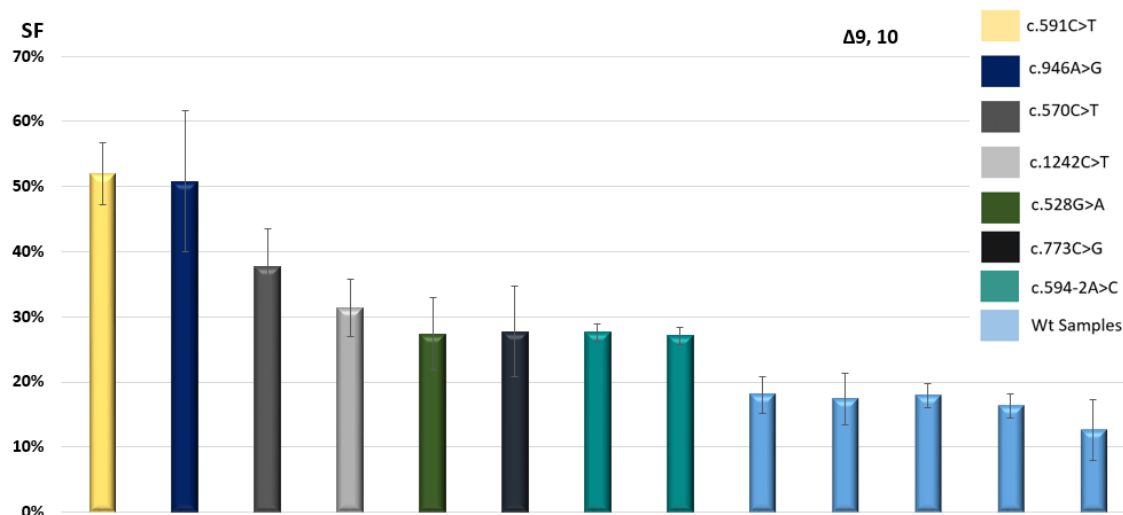
Las 6 variantes tienen en común que se localizan en el entorno de los exones 9 y 10 de *BRCA1* (exones 8, 9 y 11). Tal y como se ha comentado en la introducción (Figura 11), un estudio previo de nuestro laboratorio demostró que para la caracterización precisa por RT-PCR/EC de variantes en esta región, y en concreto la variante c.591C>T (p.Cys197=), es esencial realizar ensayos que capturen tanto el transcrito de referencia como la isoforma  $\Delta 9,10$  (Dosil et al, 2010). Por ello, realizamos en todos los casos ensayos RT-PCR/EC con cebadores localizados en los exones 8 y 11 (en el caso de la variante c.528G>A empleamos un cebador *sense* localizado en el exón 7).

En todos los casos el análisis RT-PCR/EC indica que las variantes no causan efectos cualitativos a nivel de *splicing* (se detectan los mismos transcritos en portadores y controles de sangre), aunque se observan posibles efectos cuantitativos (incremento relativo de  $\Delta 9,10$ ) especialmente en los portadores c.570C>T (p.Thr190=) y c.946A>G (p.Ser316Gly), tal y como se observa en la Figura 66. En todos los casos se confirmó expresión bi-alélica mediante secuenciación Sanger (Figura 66).



**Figura 66. Análisis del *splicing* de las variantes *BRCA1* c.528G>A (p.Thr176=), c.570C>T (p.Thr190=), c.591C>T (p.Cys197=), c.773C>G (p.Pro258Arg), c.946A>G (p.Ser316Gly) y c.1242C>T (p.Asp414=).** a) Análisis de *splicing* alternativo mediante RT/PCR-EC. Electroferograma en el que se observa la contribución relativa de los eventos  $\Delta 9,10$  (238nt),  $\Delta 10$  (284nt) y  $\Delta 9$  (315nt) y transcrito de referencia (FL, 361nt) a la expresión global detectada con el ensayo capilar E8 E11, medido en controles sanos de sangre, y en el portador de la variante c.570C>T (p.Thr190=). b) Secuenciación Sanger para estudiar expresión Bialélica de un SNP exónico en el exón 13 de *BRCA1* (c.4308C>T). El producto obtenido en una segunda PCR (cebadores E12 E14.1, sin marcar con FAM) fue secuenciado con un cebador E12 *sense*.

Para cuantificar con más precisión el efecto en los niveles de  $\Delta 9,10$  realizamos un análisis por PCR digital. Como se observa en la Figura 67 todas las variantes estudiadas parecen tener un efecto positivo en los niveles relativos de expresión de *BRCA1*  $\Delta 9,10$ . El efecto es particularmente acusado en portadores de c.591C>T y c.946A>G.

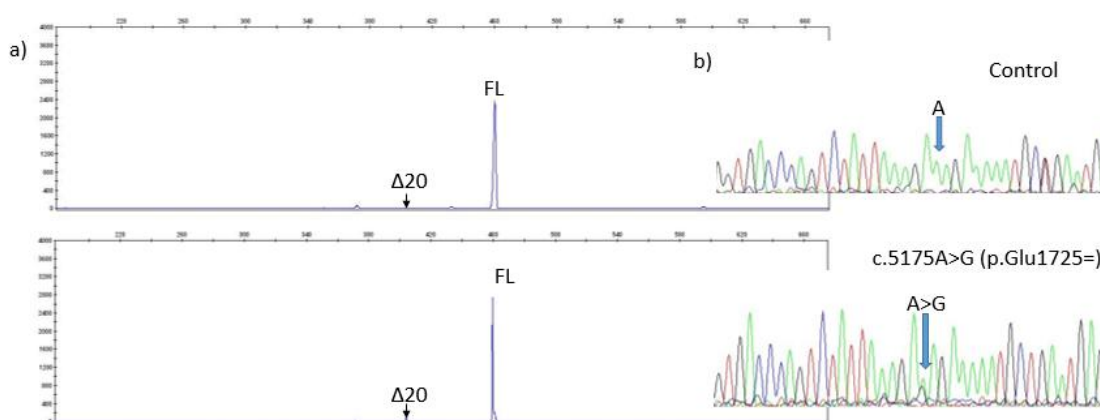


**Figura 67. Contribución relativa de  $\Delta 9,10$  a la expresión global de *BRCA1* determinada por PCR digital en portadores de las variantes *BRCA1* c.528G>A, c.570C>T, c.591C>T, c.594-2A>C, c.773C>G, c.946A>G, 1242C>T, c.528G>A y controles no portadores.** Los análisis se realizaron a partir de sangres extraídas de pacientes del HCSC (con la excepción de c.594-2A>C, que se analizó a partir de RNA extraído de una LCL). En el ensayo se utilizó una sonda FAM que reconoce la unión de los exones 8 y 11 ( $\Delta 9,10$ ) y una sonda VIC que reconoce la unión entre los exones 23 y 24 (consideramos que esta sonda captura adecuadamente la expresión global del gen *BRCA1*). La gráfica representa en porcentaje la proporción FAM/VIC, que corresponde a la denominada *splicing* fraction (SF) de *BRCA1*  $\Delta 9,10$ . La gráfica indica la precisión de cada medida determinada con el software QuantStudio 3D Analysis.

Concluimos que niveles de  $\Delta 9,10$  en torno al 50% de la expresión global de *BRCA1* no se asocian con riesgo aumentado de cáncer de mama y ovario. Por esta razón, descartamos que los niveles de *BRCA1*  $\Delta 9, 10$  observados en c.528G>A (p.Thr176=), *BRCA1* c.570C>T (p.Thr190=), *BRCA1* c.773C>G (p.Pro258Arg), *BRCA1* c.946a>G (p.Ser316Gly) y *BRCA1* c.1242C>T (p.Asp414=), aún aumentados respecto a controles, puedan asociarse con riesgo elevado de desarrollar cáncer de mama y ovario. Tras este razonamiento, consideramos que las variantes sinónimas pueden reclasificarse como Clase-1 (clasificación tras el estudio de *splicing*). En el caso de las variantes *missense*, el estudio de *splicing* no tiene ninguna repercusión sobre la clasificación inicial (podrían ser variantes patogénicas por alteración de la función de la proteína).

**BRCA1 c.5175A>G (p.Glu1725=)**

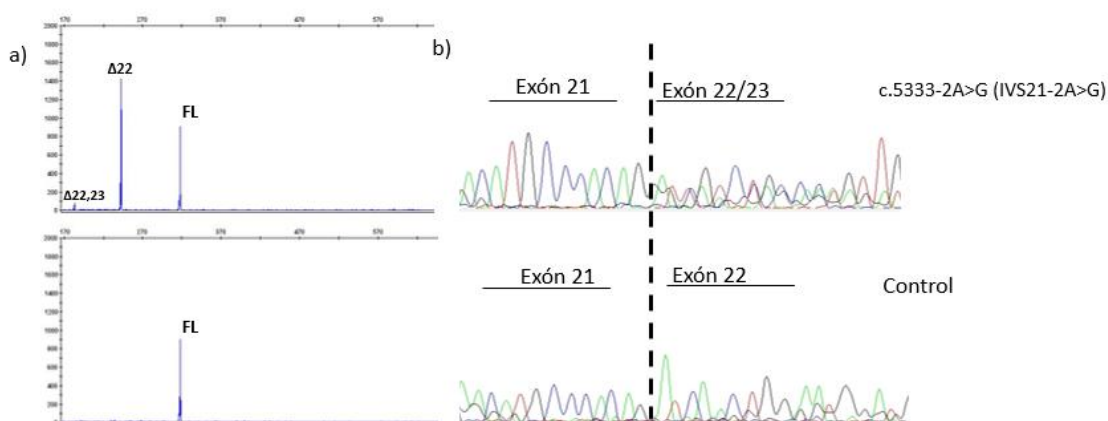
Variante sinónima localizada en el exón 19 de *BRCA1*. Un estudio mediante RT-PCR/EC (cebadores localizados en los exones 16 y 22) no detecta diferencias cualitativas ni cuantitativas entre portadores y controles (Figura 68A). El portador de la variante demuestra expresión bi-alélica (Figura 68B). En nuestra opinión, este resultado negativo reduce significativamente la probabilidad de que la variante sinónima *BRCA1* c.5175A>G (p.Glu1725=) sea patogénica (no tiene efecto funcional a nivel de proteína, y ahora sabemos que tampoco tiene efecto funcional a nivel de RNA) por lo que consideramos adecuado reclasificarla como Clase-1.



**Figura 68. Análisis del *splicing* de la variante *BRCA1* c.5175A>G (p.Glu1725=).** a) Análisis del *splicing* mediante RT-PCR/EC. Electroferograma en el que se observa la contribución relativa del evento  $\Delta 20$  (343nt), y transcrito de referencia (FL, 460nt) a la expresión global detectada con el ensayo RT-PCR E16 E22, medido en controles sanos de sangre, y en el portador de la variante c.5175A>G (p.Glu1725=). El evento  $\Delta 20$  se ha descrito previamente en controles (Colombo et al, 2014). b) Secuenciación Sanger para estudiar la expresión bialélica. El producto analizado en fragmentos, se secuenció con un cebador E16 *sense*.

**BRCA1 c.5333-2A>G (IVS21-2A>G)**

Variante localizada en el dinucleótido consenso del aceptor de *splicing* del exón 22 de *BRCA1*. El estudio mediante RT-PCR/EC (cebadores localizados en los exones 20 y 24) detecta diferencias obvias en el portador (Figura 69A). En concreto, se observa un pico muy preponderante compatible con  $\Delta 22$  (243nt) y un pico muy minoritario compatible con  $\Delta 22,23$  (183nt). El evento  $\Delta 22,23$  se ha detectado previamente en controles (Colombo et al, 2014). La identidad de  $\Delta 22$  (evento PTC-NMD) se confirmó mediante secuenciación Sanger (Figura 69B).



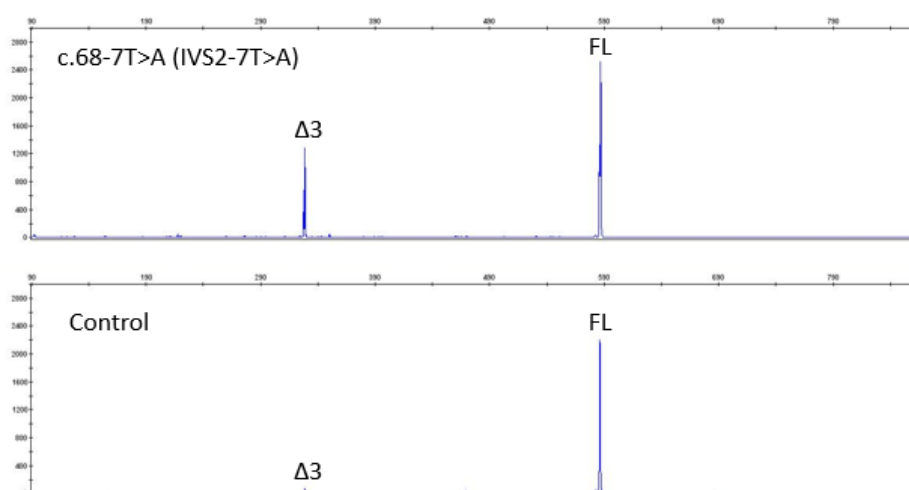
**Figura 69. Análisis del *splicing* de la variante *BRCA1* c.5333-2A>G (IVS21-2A>G).** a) Análisis de *splicing* mediante RT/PCR-EC. Electroferograma en el que se observa la contribución relativa de los eventos  $\Delta 22$  (243nt),  $\Delta 22,23$  (183nt) y transcrito de referencia (FL, 317nt), a la expresión global detectada con el ensayo RT-PCR E20 E24, medido en sangres de controles sanos y en el portador de la variante c.5333-2A>G. b) Secuenciación por Sanger. El producto de RT-PCR mostrado en el panel izquierdo se secuenció directamente con un cebador E20 *sense*. En el portador de la variante se observan dos secuencias solapantes que corresponden al exón 22 (FL) y al exón 23 ( $\Delta 22$ ).

Debido a la ausencia de posiciones exónicas heterocigotas, no hemos podido descartar experimentalmente la producción de transcritos de referencia por parte del alelo c.5333-2G. Sin embargo, dada la localización de la variante (sitio consenso de *splicing*) y la proporción de transcrito aberrante  $\Delta 22$  sugerida por la electroforesis capilar y la secuenciación Sanger (aprox. 50%), consideramos adecuado reclasificar la variante como Clase-5.

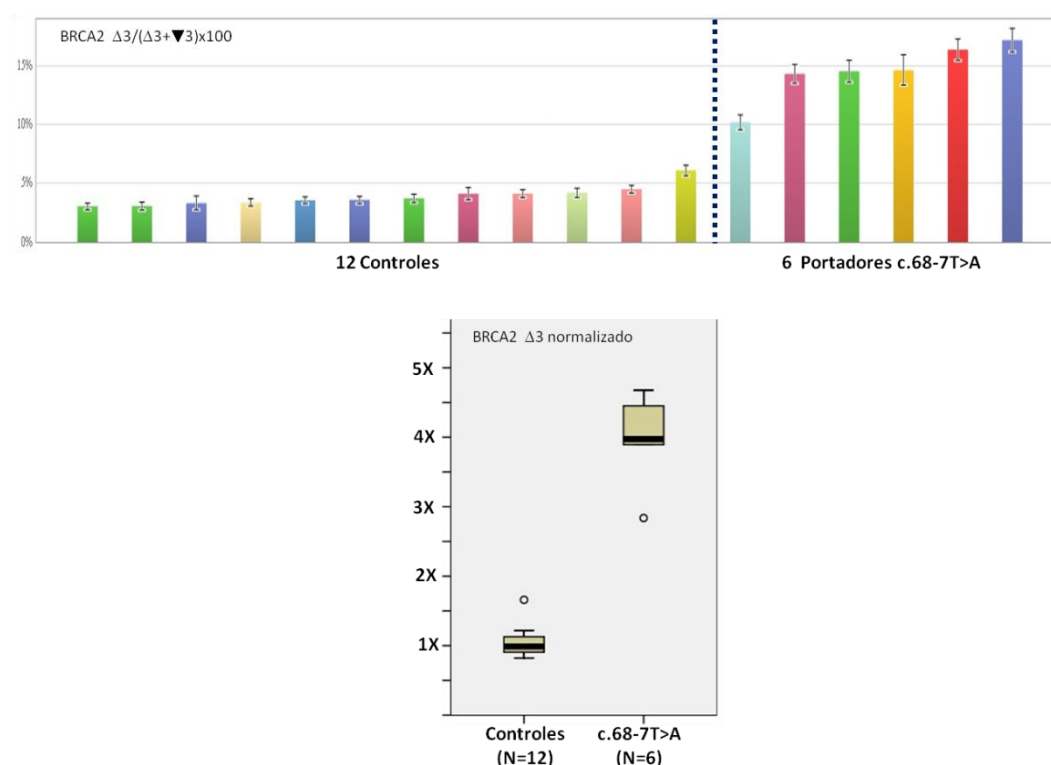
## 2. Análisis de *splicing* en portadores de variantes genéticas en *BRCA2*

### ***BRCA2* c.68-7T>A (IVS2-7T>A)**

Numerosos estudios previos (algunos realizados a partir de RNA de LCLs de portadores, otros en minigenes) coinciden en describir un aumento relativo de la isoforma  $\Delta 3$  asociado a esta variante (Santarosa et al, 1999; Vreeswijk et al, 2009; Sanz et al, 2010, Muller et al, 2011; Houdayer et al, 2012). Nosotros hemos reproducido esta observación mediante RT-PCR/EC (cebadores localizados en los exones 1 y 6), tal y como se muestra en la Figura 70. Un estudio reciente llevado a cabo por los consorcios internacionales ENIGMA, CIMBA y BCAC en el que hemos participado (manuscrito en revisión) ha demostrado mediante análisis multifactorial que la variante *BRCA2* c.68-7T>A (IVS2-7T>A) no es patogénica (Clase-1). Por ello, decidimos cuantificar de forma precisa la contribución de  $\Delta 3$  a la expresión global de *BRCA2* en portadores mediante PCR digital (Figura 71).



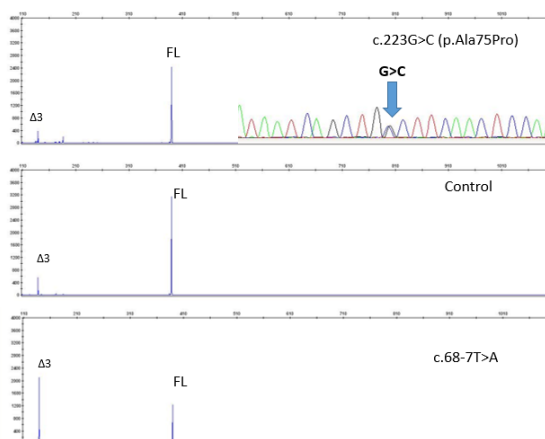
**Figura 70. Análisis del *splicing* de la variante *BRCA2* c.68-7T>A (IVS2-7T>A) mediante RT-PCR/EC. a) Análisis de *splicing* mediante RT/PCR-EC.** Electroferograma en el que se observa la contribución relativa de  $\Delta 3$  (327nt) y el transcrito de referencia (FL, 576nt) a la expresión global detectada con cebadores en los exones 1 y 6. El experimento se realizó a partir de RNA obtenido de LCLs de un portador de c.68-7T>A y un control negativo. EL experimento fue realizado en condiciones semi-cuantitativas (33 ciclos). Sin embargo, con toda probabilidad, el nivel de  $\Delta 3$  está sobrestimado debido a amplificación preferencial.



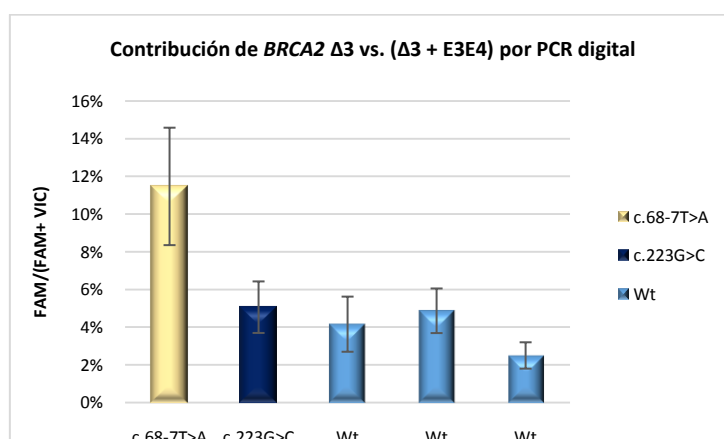
**Figura 71. Contribución relativa de  $\Delta 3$  a la expresión global de *BRCA2* determinada por PCR digital.** Los experimentos se realizaron a partir de RNA extraído de 6 LCLs portadoras de la variante *BRCA2* c.68-2T>A y 12 LCLs control no portadoras. Se utilizó un ensayo TaqMan FAM que reconoce la unión de los exones 2 y 4 de *BRCA2* ( $\Delta 3$ ) y un ensayo TaqMan VIC que reconoce la unión de los exones 3 y 4; y por tanto, el transcrito de referencia que contiene el exón 3 ( $\nabla 3$ ). La gráfica superior muestra (ordenados de menor a mayor) el resultado de las 18 determinaciones individuales (contribución de  $\Delta 3$  a la expresión global  $\Delta 3 + \nabla 3$ ) calculado por el software QuantStudio 3D Analysis Cloud (el software también calcula de forma automática la precisión de cada medida). En el panel inferior se muestran los resultados del panel superior, agrupados por genotipo y normalizados respecto al valor medio en controles. Los diagramas de cajas muestran el valor mínimo, percentil 25%, mediana, percentil 75% y valor máximo (también se destacan dos valores extremos). La contribución media de  $\Delta 3$  a la expresión global de *BRCA2* en portadores *BRCA2* c.68-2T>A (13,4%; rango 10,5%-17,2%) es cuatro veces superior a la observada en controles (3,7%; rango 3,0%-6,1%;  $p < 10^{-4}$ ).

**BRCA2 c.223G>C (p.Ala75Pro)**

Variante *missense* localizada en el exón 3 de *BRCA2*. Un análisis multifactorial (Easton et al, 2007) demostró hace tiempo que la variante no es patogénica (Clase-1). Hemos realizado un estudio RT-PCR/EC (cebadores localizados en los exones 2 y 4) que sugiere que la variante no modifica los niveles de *BRCA2*  $\Delta 3$  respecto a controles (Figura 72). Este resultado negativo fue confirmado por PCR digital (Figura 73).



**Figura 72. Análisis del *splicing* de la variante *BRCA2* c.223G>C (p.Ala75Pro) mediante RT/PCR-EC. a) Análisis de *splicing* mediante RT/PCR-EC.** Electroferograma en el que se observa la contribución relativa del evento  $\Delta 3$  (138nt) y el transcrito de referencia (FL, 390nt), a la expresión global detectada con cebadores localizados en los exones 2 y 4. El experimento se realizó a partir de RNA obtenido de sangre de un portador de la variante c.223G>C (p. Ala75Pro), controles negativos de sangre, y un control positivo de LCL (c.68-7T>A). **b) Secuenciación Sanger** del producto obtenido en una segunda RT-PCR (entre los exones 2 y 6) utilizando un cebador E2 *sense*.

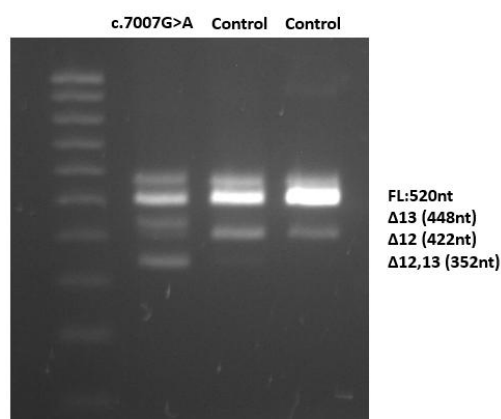


**Figura 73. Contribución relativa de  $\Delta 3$  a la expresión global de *BRCA2* determinada por PCR digital en portadores de las variantes *BRCA2* c.223G>C, 3 controles negativos (individuos controles sanos) y un control positivo c.68-7T>A.** Los análisis se realizaron a partir de RNA de sangre de un portador de la variante *BRCA2* c.223G>C, controles sanos de sangre (control negativo), y un portador LCL de la variante *BRCA2* c.68-7T>A (control positivo). En el ensayo se utilizó una sonda FAM que reconoce la unión de los exones 2 y 4 ( $\Delta 3$ ) y una sonda VIC que reconoce la unión entre los exones 3 y 4 ( $\nabla 3$ ). La gráfica representa en porcentaje la proporción FAM/(FAM+VIC) que corresponde a la contribución de  $\Delta 3$  a la expresión global de *BRCA2*. Se indica la precisión de cada medida determinada con el software QuantStudio 3D Analysis.

**BRCA2 c.7007G>A (p.Arg2336His)**

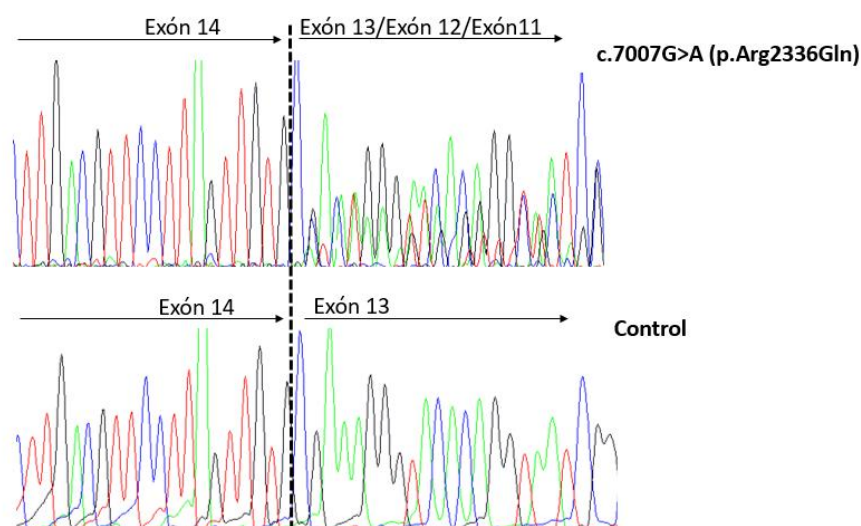
Variante *missense* localizada en el último nucleótido del exón 13 de *BRCA2*. La variante está anotada en BRCAShare como Clase-5, pero en ClinVar tiene varias anotaciones conflictivas. Se han realizado numerosos estudios de *splicing* que coinciden en señalar que la variante afecta al *splicing*. Sin embargo, según qué estudio la variante genera: transcritos  $\Delta 13$  (Thomanssen et al, 2006; Sanz et al, 2010), transcritos  $\Delta 12,13$  (Houdayer C et al., 2012), o ambos tipos de transcritos (Machackova et al, 2008). Tanto  $\Delta 13$  como  $\Delta 12,13$  son transcritos PTC-NMD.

Nosotros hemos realizado un estudio por RT-PCR seguido de análisis en agarosa y secuenciación Sanger que detecta ambos transcritos (Figuras 74 y 75).



**Figura 74. Análisis del *splicing* de la variante *BRCA2* c.7007G>A (p.Arg2336His) mediante RT-PCR/Agarosa.** El experimento se realizó en Gel al 2% en agarosa en el que se observan los eventos:  $\Delta 12$  (422nt),  $\Delta 13$  (448nt) y  $\Delta 12,13$  (352nt); además del transcrito de referencia (510nt) obtenidos en una PCR entre los exones 11 (B2 E11s) y 14 (B2 E14.1a). El estudio se ha realizado en RNAs extraídos de sangre periférica de controles sanos y en el portador de la variante c.7007G>A (p.Arg2336Gln).

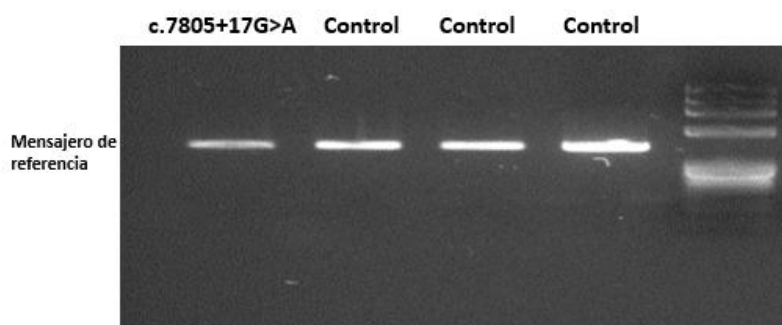




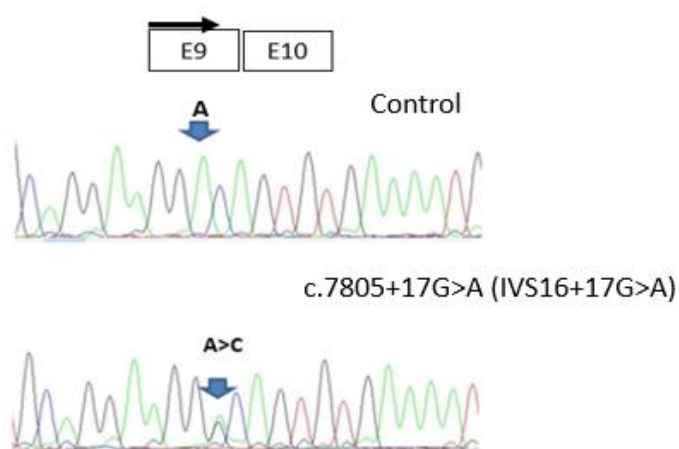
**Figura 75. Análisis del *splicing* de la variante *BRCA2* c.7007G>A (p.Arg2336His) mediante secuenciación Sanger.** El producto RT-PCR E11-E14 mostrado en la Figura 74 se secuenció directamente con un cebador *antisense* en el exón 14 (B2 E14.1a). En el portador de la variante se observa un solapamiento de 3 secuencias que corresponde a los exones 13 (transcrito de referencia), 12 ( $\Delta 13$ ) y 11 ( $\Delta 12,13$ ).

**BRCA2 c.7805+17G>A**

La variante se localiza en el intrón 16 del gen *BRCA2*. Los análisis resumidos en las figuras 76 y 77 demuestran que la variante no afecta al *splicing*.



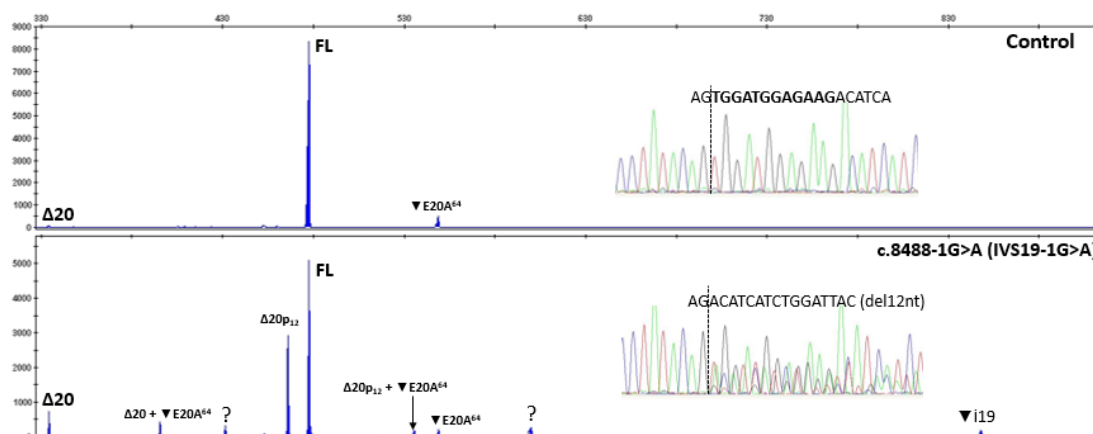
**Figura 76. Análisis del *splicing* de la variante *BRCA2* c.7805+17G>A mediante RT-PCR /Agarosa. Análisis de *splicing* mediante RT-PCR/Agarosa.** Gel al 2% en agarosa en el que se observa el transcrito de referencia como único producto, detectado en un ensayo RT-PCR E15-E17.1. El estudio se ha realizado en controles de sangre y en el portador de la variante c.7805+17G>A. El producto de RT-PCR mostrado en la figura se secuenció con un cebador *sense* para confirmar que no hay ninguna alteración de *splicing* en la muestra. Para confirmar expresión bi-alélica en la muestra, se realizó el experimento mostrado en la Figura 77.



**Figura 77. Análisis del *splicing* de la variante *BRCA2* c.7805+17G>A mediante Secuenciación Sanger. Secuenciación Sanger para analizar la expresión alélica en el SNP c.865A>C (rs766173 A/C).** El producto obtenido de una segunda RT-PCR con cebadores en los exones 9 (B2 E9s) y 10 (B2 E10.2a) se secuenció con el cebador *sense* (B2 E9s). La presencia del cambio *BRCA2* c.865A>C (rs766173) demostró expresión bi-alélica en el portador.

**BRCA2 c.8488-1G>A (IVS19-1G>A)**

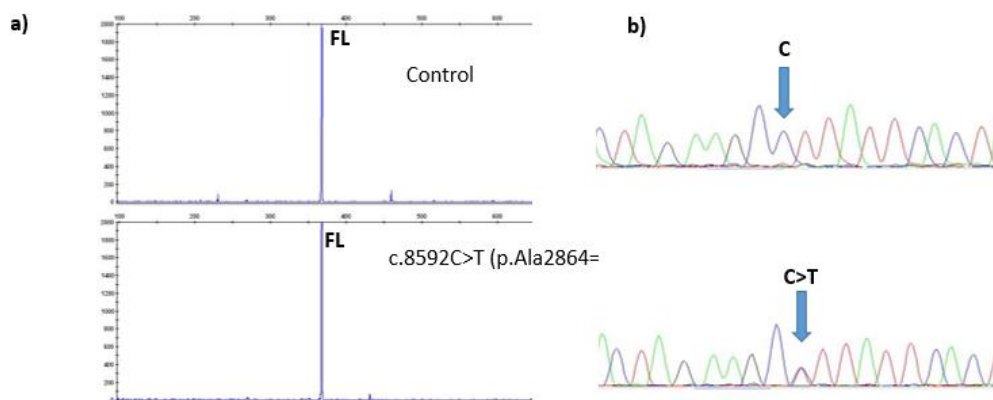
Variante intrónica situada en el aceptor de *splicing* del exón 20 de *BRCA2*. Se han realizado estudios previos para valorar el efecto de esta variante en el *splicing* (Acedo et al, 2012; Santos et al, 2014, Acedo et al, 2015). Estos estudios ofrecen resultados discordantes (ver discusión) por lo que nuestro estudio (Figura 78), pretende ser aclaratorio.



**Figura 78. Análisis del *splicing* de la variante *BRCA2* c.8488-1G>A (IVS19-1G>A). Análisis de *splicing* mediante RT/PCR-EC y secuenciación Sanger.** Los experimentos se realizaron a partir de RNA extraído de sangre periférica de un portador de la variante *BRCA2* c.8488-1G>A (IVS19-1G>A) y un control sano. En el electroferograma se observa la contribución relativa de los eventos  $\Delta 20$  (334nt),  $\Delta 20 + \nabla E20A^{64}$  (395nt),  $\Delta 20p_{12}$  (466nt), FL (478nt),  $\Delta 20p_{12} + \nabla E20A^{64}$  (535nt),  $\nabla E20A^{64}$  (548nt) y  $\nabla i19$  (847nt) a la expresión global detectada con el ensayo RT-PCR E18-E22. También se detectan picos minoritarios que no hemos sabido anotar (?). El uso de un estándar de tamaños LIZ-500 no permite medir con precisión el tamaño de los picos >500nt. Los eventos  $\Delta 20$  y  $\nabla E20A^{64}$  se han descrito previamente como eventos de *splicing* alternativo muy minoritarios en controles (Fackenthal et al, 2016). El producto de la RT-PCR se secuenció con un cebador *sense* en el exón 18 (B2 E18s). En el portador se observa un exón 19 limpio seguido de un solapamiento de 3 secuencias que corresponden, en orden de intensidad, al exón 20 (FL),  $\Delta 20p_{12}$ , y retención del intrón 19 ( $\nabla i19$ ). Los resultados mostrados sugieren que el alelo *BRCA2* c.8488-1A produce mayoritariamente  $\Delta 20p_{12}$ , pero también niveles bajos de  $\nabla i19$  y  $\Delta 20$ . Experimentos realizados con una RT-PCR E19-E22 reprodujo los resultados.

**BRCA2 c.8592C>T (p.Ala2864=)**

Variante sinónima localizada en el exón 20 de *BRCA2*. No hemos identificado en la literatura científica estudios previos de *splicing* para esta variante. Los experimentos mostrados en la Figura 79 demuestran que la variante no afecta al *splicing*.



**Figura 79. Análisis del *splicing* de la variante *BRCA2* c.8592C>T (p.Ala2864=).** a) Análisis de *splicing* mediante RT-PCR/EC. El ensayo, realizado con los cebadores E19-E22, sólo detecta el transcrito de referencia (FL, 365nt). Obsérvese que el pico de 365nt está completamente saturado. Los experimentos se realizaron a partir de RNA extraído de sangre periférica de un control sano y un portador de la variante c.8592C>T. b) Secuenciación Sanger. El producto RT-PCR mostrado en el panel izquierdo se secuenció con un cebador *sense* en el exón 19 (B2 E19s), demostrándose expresión bi-alélica.



# Referencias



# REFERENCIAS

---

1. Acedo A, Hernández-Moro C, Curiel-García Á, Díez-Gómez B, Velasco EA. Functional Classification of *BRCA2* DNA Variants by Splicing Assays in a Large Minigene with 9 Exons. *Human Mutation*. 2015. 36: 210-21.
2. Acedo A, Sanz DJ, Durán M, Infante M, Pérez-Cabornero L, Miner C, et al. Comprehensive splicing functional analysis of DNA variants of the *BRCA2* gene by hybrid minigenes. *Breast Cancer Res*. 2012. 14: R87.
3. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. *Nature*. 2013. 500: 415-21.
4. Antoniou AC, Casadei S, Heikkinen T, Barrowdale D, Pyrkäs K, Roberts J, et al. Breast-Cancer Risk in Families with Mutations in *PALB2*. *New England Journal of Medicine*. 2014. 371: 497-506.
5. Baer R, Ludwig T. The *BRCA1/BARD1* heterodimer, a tumor suppressor complex with ubiquitin E3 ligase activity. *Curr Opin Genet Dev*. 2002. 12: 86-91.
6. Bahcall O. Polygenic architecture. *Nature Genetics*. 2013. 45: 1417-1417.
7. Bignell G, Micklem G, Stratton MR, Ashworth A, Wooster R. The BRC repeats are conserved in mammalian *BRCA2* proteins. *Hum Mol Genet*. 1997. 6: 53-8.
8. Brohl AS, Patidar R, Turner CE, Wen X, Song YK, Wei JS, et al. Frequent inactivating germline mutations in DNA repair genes in patients with Ewing sarcoma. *Genet Med*. 2017. 19: 955-8.
9. Buisson R, Masson J-Y. *PALB2* self-interaction controls homologous recombination. *Nucleic Acids Research*. 2012. 40: 10312-23.
10. Buisson R, Niraj J, Pauty J, Maity R, Zhao W, Coulombe Y, Sung P and Masson J-Y. Breast cancer proteins *PALB2* and *BRCA2* stimulate polymerase  $\eta$  in recombination-associated DNA synthesis at blocked replication forks. *Cell rep*. 2014. 13: 553-564.
11. Casadei S, Norquist BM, Walsh T, Stray S, Mandell JB, Lee MK, Stamatoyannopoulos JA, King MC. Contribution of inherited mutations in the *BRCA2*-interacting protein *PALB2* to familial Breast Cancer. *Cancer Res*. 2011. 71: 2222-9.
12. Catucci I, Peterlongo P, Ciceri S, Colombo M, Pasquini G, Barile M, Bonanni B, Verderio P, Pizzamiglio S, Foglia C, Falanga A, Marchetti M, et al. *PALB2* sequencing in Italian familial breast cancer cases reveals a high-risk mutation recurrent in the province of Bergamo. *Genet. Med*. 2014. 16: 688-694.



13. Ceccaldi R, Sarangi P, D'Andrea AD. The Fanconi anaemia pathway: new players and new functions. *Nat Rev Mol Cell Biol.* 2016. 17: 337-49.
14. Chun J, Buechelmaier ES, Powell SN. Rad51 paralogs BCDX2 and CX3 act at different stages in the BRCA1-BRCA2-dependent homologous recombination pathway. *Mol Cell Biol.* 2013. 33: 387-95.
15. Clark SL, Rodriguez AM, Snyder RR, Hankins GDV, Boehning D. STRUCTURE-FUNCTION OF THE TUMOR SUPPRESSOR BRCA1. *Computational and Structural Biotechnology Journal.* 2012. 1.
16. Colombo M, Blok MJ, Whiley P, Santamariña M, Gutiérrez-Enríquez S, Romero A, et al. Comprehensive annotation of splice junctions supports pervasive alternative splicing at the BRCA1 locus: a report from the ENIGMA consortium. *Hum Mol Genet.* 2014. 23: 3666-80.
17. Couch FJ, Nathanson KL, Offit K. Two decades after BRCA: setting paradigms in personalized cancer care and prevention. *Science.* 2014. 343: 1466-70.
18. Couch FJ, Wang X, McGuffog L, Lee A, Olswold C, Kuchenbaecker KB, et al. Genome-Wide Association Study in BRCA1 Mutation Carriers Identifies Novel Loci Associated with Breast and Ovarian Cancer Risk. Hunter KW. *PLoS Genetics.* 2013. 9.
19. Coulet F, Fajac A, Colas C, Eyries M, Dion-Minière A, Rouzier R, et al. Germline *RAD51C* mutations in ovarian cancer susceptibility: Ovarian cancer and germline *RAD51C* mutations. *Clinical Genetics.* 2013. 83: 332-6.
20. D'Andrea AD, Grompe M. The Fanconi anaemia/BRCA pathway. *Nat Rev Cancer.* 2003. 3: 23-34.
21. D'Andrea AD. Susceptibility pathways in Fanconi's anemia and breast cancer. *N Engl J Med.* 2010. 362: 1909-19.
22. De la Hoya M, Soukarieh O, **López-Perolio I**, Vega A, Walker LC, van Ierland Y, et al. Combined genetic and splicing analysis of BRCA1 c.[594-2A>C; 641A>G] highlights the relevance of naturally occurring in-frame transcripts for developing disease gene variant classification algorithms. *Hum Mol Genet.* 2016. 25: 2256-68.
23. Desmond A, Kurian AW, Gabree M, Mills MA, Anderson MJ, Kobayashi Y, et al. Clinical Actionability of Multigene Panel Testing for Hereditary Breast and Ovarian Cancer Risk Assessment. *JAMA Oncology.* 2015. 1: 943.
24. Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, et al. Landscape of transcription in human cells. *Nature.* 2012. 489: 101-8.
25. Doblin A et al. STAR: ultrafast universal RNA-seq aligner. 2014. Oxford Academic. 29: 15-21.

26. Dosanjh MK, Collins DW, Fan W, Lennon GG, Albala JS, Shen Z, et al. Isolation and characterization of RAD51C, a new human member of the RAD51 family of related genes. *Nucleic Acids Res.* 1998. 26: 1179-84.
27. Dosil V, Tosar A, Canadas C, Perez-Segura P, Diaz-Rubio E, Caldes T, et al. Alternative Splicing and Molecular Characterization of Splice Site Variants: BRCA1 c.591C>T; as a Case Study. *Clinical Chemistry.* 1 de enero de 2010. 56: 53-61.
28. Easton DF, Pharoah PDP, Antoniou AC, Tischkowitz M, Tavtigian SV, Nathanson KL, et al. Gene-Panel Sequencing and the Prediction of Breast-Cancer Risk. *New England Journal of Medicine.* 2015. 372: 2243-57.
29. Eccles DM, Mitchell G, Monteiro ANA, Schmutzler R, Couch FJ, Spurdle AB, et al. BRCA1 and BRCA2 genetic testing-pitfalls and recommendations for managing variants of uncertain clinical significance. *Ann Oncol.* 2015. 26: 2057-65.
30. ElShamy, W. M., Livingston, D. M. Identification of BRCA1-IRIS, a BRCA1 locus product. *Nature Cell Biol.* 2004. 6: 954-967.
31. Fackenthal JD, Yoshimatsu T, Zhang B, de Garibay GR, Colombo M, De Vecchi G, Ayoub SC, Lal K, Olopade OI, Vega A, Santamariña M, Blanco A, Wappenschmidt B, Becker A, Houdayer C, Walker LC, **López-Perolio I**, Thomassen M, Parsons M, Whiley P, Blok MJ, Brandão RD, Tserpelis D, Baralle D, Montalban G, Gutiérrez-Enríquez S, Díez O, Lazaro C; kConFaB Investigators, Spurdle AB, Radice P, de la Hoya M. Naturally occurring *BRCA2* alternative mRNA *splicing* events in clinically relevant samples. *J Med Genet.* 2016 53: 548-58.
32. Farrugia DJ, Agarwal MK, Pankratz VS, Deffenbaugh AM, Pruss D, Frye C, et al. Functional assays for classification of BRCA2 variants of uncertain significance. *Cancer Res.* 2008. 68: 3523-31.
33. Foulkes WD, Ghadirian P, Akbari MR, Hamel N, Giroux S, Sabbaghian N, et al. Identification of a novel truncating PALB2 mutation and analysis of its contribution to early-onset breast cancer in French-Canadian women. *Breast Cancer Research.* 2007. 9.
34. Fortin J, Moisan AM, Dumont M, Leblanc G, Labrie Y, Durocher F *et al.* *Biochim Biophys Acta.* 2005. 1731: 57–65.
35. Fraile-Bethencourt E, Díez-Gómez B, Velásquez-Zapata V, Acedo A, Sanz DJ, Velasco EA. Functional classification of DNA variants by hybrid minigenes: Identification of 30 spliceogenic variants of BRCA2 exons 17 and 18. *PLoS Genet.* 2017. 13.
36. Fox EJ, Salk JJ, Loeb LA. Exploring the implications of distinct mutational signatures and mutation rates in aging and cancer. *Genome Med.* 8.

37. Friebel TM, Domchek SM, Rebbeck TR. Modifiers of Cancer Risk in BRCA1 and BRCA2 Mutation Carriers: A Systematic Review and Meta-Analysis. *JNCI: Journal of the National Cancer Institute*. 2014. 106.
38. Ghoussaini M, Pharoah PDP, Easton DF. Inherited genetic susceptibility to breast cancer: the beginning of the end or the end of the beginning? *Am J Pathol*. 2013. 183: 1038-51.
39. Gilbert W. Why genes in pieces? *Nature*. 1978. 271: 501.
40. Goldgar DE, Easton DF, Deffenbaugh AM, Monteiro ANA, Tavtigian SV, Couch FJ. Integrated Evaluation of DNA Sequence Variants of Unknown Clinical Significance: Application to BRCA1 and BRCA2. *The American Journal of Human Genetics*. 2004. 75: 535-44.
41. Golmard L, et al. Germline mutation in the RAD51B gene confers predisposition to Breast cancer. *BMC Cancer*. 2013. 19: 13-484.
42. Greenman C, Stephens P, Smith R, Dalgliesh GL, Hunter C, Bignell G, et al. Patterns of somatic mutation in human cancer genomes. *Nature*. 2007. 446: 153-8.
43. Gutiérrez-Enríquez S, Bonache S, Ruíz de Garibay G, Osorio A, Santamariña M, Ramón y Cajal T, et al. About 1% of the breast and ovarian Spanish families testing negative for *BRCA1* and *BRCA2* are carriers of *RAD51D* pathogenic variants: *RAD51D* germline mutations in breast and ovarian Spanish families. *International Journal of Cancer*. 2014. 134: 2088-97.
44. Hanahan D, Weinberg RA. Hallmarks of Cancer: The Next Generation. *Cell*. 2011. 144: 646-74.
45. Hegele A, Kamburov A, Grossmann A, Sourlis C, Wowro S, Weimann M, et al. Dynamic Protein-Protein Interaction Wiring of the Human Spliceosome. *Molecular Cell*. 2012. 45: 567-80.
46. Holloman WK. Unraveling the mechanism of BRCA2 in homologous recombination. *Nature Structural & Molecular Biology*. 2011. 18: 748-54.
47. Houdayer C, Caux-Moncoutier V, Krieger S, Barrois M, Bonnet F, Bourdon V, et al. Guidelines for splicing analysis in molecular diagnosis derived from a set of 327 combined in silico/in vitro studies on BRCA1 and BRCA2 variants. *Hum Mutat*. 2012. 33: 1228-38.
48. Howlett NG, Taniguchi T, Olson S, Cox B, Waisfisz Q, De Die-Smulders C, et al. Biallelic inactivation of BRCA2 in Fanconi anemia. *Science*. 2002. 297: 606-9.
49. Janatova M, Soukupova J, Stribrna J, Kleiblova P, Vocka M, Boudova P, et al. Mutation Analysis of the RAD51C and RAD51D Genes in High-Risk Ovarian Cancer

- Patients and Families from the Czech Republic. Brusgaard K, editor. PLOS ONE. 2015. 10.
50. Jensen RB, Carreira A, Kowalczykowski SC. Purified human BRCA2 stimulates RAD51-mediated recombination. *Nature*. 2010. 467: 678-83.
51. Jervis S, Song H, Lee A, Dicks E, Tyrer J, Harrington P, et al. Ovarian cancer familial relative risks by tumour subtypes and by known ovarian cancer genetic susceptibility variants. *Journal of Medical Genetics*. 2014. 51: 108-13.
52. Joenje H, Patel KJ. The emerging genetic and molecular basis of Fanconi anaemia. *Nat Rev Genet*. 2001. 2: 446-57.
53. Kiiski JI, Pelttari LM, Khan S, Freysteinsdottir ES, Reynisdottir I, Hart SN, et al. Exome sequencing identifies FANCM as a susceptibility gene for triple-negative breast cancer. *Proc Natl Acad Sci USA*. 2014. 111: 15172-7.
54. Kim YM and Choi BS. Structural and functional characterization of N-terminal domain of human RAD51D. *Int J Biochem Cell Biol*. 2011. 43: 416-22.
55. Konstantinopoulos PA, Ceccaldi R, Shapiro GI, D'Andrea AD. Homologous Recombination Deficiency: Exploiting the Fundamental Vulnerability of Ovarian Cancer. *Cancer Discov*. 2015. 5: 1137-54.
56. Korde LA, Mueller CM, Loud JT, Struewing JP, Nichols K, Greene MH, et al. No evidence of excess breast cancer risk among mutation-negative women from BRCA mutation-positive families. *Breast Cancer Research and Treatment*. 2011. 125: 169-73.
57. Kraus C, Hoyer J, Vasileiou G, Wunderle M, Lux MP, Fasching PA, et al. Gene panel sequencing in familial breast/ovarian cancer patients identifies multiple novel mutations also in genes others than BRCA1/2: Gene panel sequencing in familial breast/ovarian cancer patients [AQ]. *International Journal of Cancer*. 1 2017. 140: 95-102.
58. Kurian AW, Hare EE, Mills MA, Kingham KE, McPherson L, Whittemore AS, et al. Clinical Evaluation of a Multiple-Gene Sequencing Panel for Hereditary Cancer Risk Assessment. *Journal of Clinical Oncology*. 2014. 32: 2001-9.
59. Kurian AW, Ford JM. Multigene Panel Testing in Oncology Practice: How Should We Respond? *JAMA Oncology*. 2015. 1: 277.
60. Levy-Lahad E. Fanconi anemia and breast cancer susceptibility meet again. *Nature Genetics*. 2010. 42: 368-9.
61. Li J, Meeks H, Feng B-J, Healey S, Thorne H, Makunin I, et al. Targeted massively parallel sequencing of a panel of putative breast cancer susceptibility genes in a large cohort of multiple-case breast and ovarian cancer families. *Journal of Medical Genetics*. 2016. 53: 34-42.

62. Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, et al. Environmental and Heritable Factors in the Causation of Cancer Analyses of Cohorts of Twins from Sweden, Denmark, and Finland. *New England Journal of Medicine*. 2000. 343: 78-85.
63. Lin Z, Kong H, Nei M, Ma H. Origins and evolution of the recA/RAD51 gene family: evidence for ancient gene duplication and endosymbiotic gene transfer. *Proc Natl Acad Sci USA*. 2006. 103: 10328-33.
64. Lindor NM, Guidugli L, Wang X, Vallée MP, Monteiro ANA, Tavtigian S, et al. A review of a multifactorial probability-based model for classification of BRCA1 and BRCA2 variants of uncertain significance (VUS). *Human Mutation*. 2012. 33: 8-21.
65. Loveday C, Turnbull C, Ramsay E, Hughes D, Ruark E, Frankum JR, et al. Germline mutations in RAD51D confer susceptibility to ovarian cancer. *Nature Genetics*. 2011. 43: 879-82.
66. Lu C, Xie M, Wendl MC, Wang J, McLellan MD, Leiserson MDM, et al. Patterns and functional implications of rare germline variants across 12 cancer types. *Nature Communications*. 2015. 6:10086.
67. Lu Z-X, Jiang P, Xing Y. Genetic variation of pre-mRNA alternative splicing in human populations. *Wiley Interdisciplinary Reviews: RNA*. 2011. 3: 581-92.
68. Ludwig T, Chapman DL, Papaioannou VE, Efstratiadis A. Targeted mutations of breast cancer susceptibility gene homologs in mice: lethal phenotypes of Brca1, Brca2, Brca1/Brca2, Brca1/p53, and Brca2/p53 nullizygous embryos. *Genes Dev*. 1997. 11: 1226-41.
69. Machackova E, Foretova L, Lukesova M, Vasickova P, Navratilova M, Coene I, et al. Spectrum and characterisation of BRCA1 and BRCA2 deleterious mutations in high-risk Czech patients with breast and/or ovarian cancer. *BMC Cancer*. 2008. 8: 140.
70. Mavaddat N, Peock S, Frost D, Ellis S, Platte R, Fineberg E, et al. Cancer Risks for BRCA1 and BRCA2 Mutation Carriers: Results From Prospective Analysis of EMBRACE. *JNCI: Journal of the National Cancer Institute*. 2013. 105: 812-22.
71. Mavaddat N, Pharoah PDP, Michailidou K, Tyrer J, Brook MN, Bolla MK, et al. Prediction of Breast Cancer Risk Based on Profiling With Common Genetic Variants. *JNCI: Journal of the National Cancer Institute*. 2015. 107: 5.
72. McCarthy AM, Armstrong K. The Role of Testing for *BRCA1* and *BRCA2* Mutations in Cancer Prevention. *JAMA Internal Medicine*. 2014. 174: 1023.
73. Meeks HD, Song H, Michailidou K, Bolla MK, Dennis J, Wang Q, et al. BRCA2 Polymorphic Stop Codon K3326X and the Risk of Breast, Prostate, and Ovarian Cancers. *J Natl Cancer Inst*. 2016. 108.

74. Meindl A, Hellebrand H, Wiek C, Erven V, Wappenschmidt B, Niederacher D, et al. Germline mutations in breast and ovarian cancer pedigrees establish RAD51C as a human cancer susceptibility gene. *Nature Genetics*. 2010. 42: 410-4.
75. Mercer TR, Gerhardt DJ, Dinger ME, Crawford J, Trapnell C, Jeddloh JA, et al. Targeted RNA sequencing reveals the deep complexity of the human transcriptome. *Nature Biotechnology*. 2011. 30: 99-104.
76. Miller KA. Domain mapping of the Rad51 paralogs protein complexes. *Nucleic Acids Research*. 2004. 32: 169-78.
77. Morgan WF, Corcoran J, Hartmann A, Kaplan MI, Limoli CL, Ponnaiya B. DNA double-strand breaks, chromosomal rearrangements, and genomic instability. *Mutat Res*. 1998. 404: 125-8.
78. Mudge JM, Frankish A, Fernandez-Banet J, Alioto T, Derrien T, Howald C, et al. The Origins, Evolution, and Functional Potential of Alternative Splicing in Vertebrates. *Molecular Biology and Evolution*. 2011. 28: 2949-59.
79. Mudge JM, Frankish A, Harrow J. Functional transcriptomics in the post-ENCODE era. *Genome Res*. 2013. 23: 1961-73.
80. Muller D, Rouleau E, Schultz I, Caputo S, Lefol C, Bièche I, et al. An entire exon 3 germ-line rearrangement in the BRCA2 gene: pathogenic relevance of exon 3 deletion in breast cancer predisposition. *BMC Med Genet*. 2011. 12:121.
81. Narod SA, Feunteun J, Lynch HT, Watson P, Conway T, Lynch J, et al. Familial breast-ovarian cancer locus on chromosome 17q12-q23. *Lancet*. 1991. 338: 82-3.
82. Neidhardt G, Hauke J, Ramser J, Groß E, et al. Association between Loss of function mutations within the FANCM gene and early-onset familial Breast cancer. *JAMA Oncol*. 2017. 1: 1245-1248.
83. Norquist BM, Harrell MI, Brady MF, Walsh T, Lee MK, Gulsuner S, et al. Inherited Mutations in Women With Ovarian Carcinoma. *JAMA Oncology*. 1 de abril de 2016. 2: 482.
84. Orr KS, Savage KI. The BRCA1 and BRCA2 Breast and Ovarian Cancer Susceptibility Genes Implications for DNA Damage Response, DNA Repair and Cancer Therapy. En: Chen CC, editor. *Advances in DNA Repair*. 2015.
85. Osorio A, Endt D, Fernandez F, Eirich K, de la Hoya M, Schmutzler R, et al. Predominance of pathogenic missense variants in the RAD51C gene occurring in breast and ovarian cancer families. *Human Molecular Genetics*. 2012. 21: 2889-98.
86. Park J-Y, Zhang F, Andreassen PR. PALB2: the hub of a network of tumor suppressors involved in DNA damage responses. *Biochim Biophys Acta*. 2014; 1846: 263-75.

87. Pittman DL, Weinberg LR, Schimenti JC. Identification, Characterization, and Genetic Mapping of Rad51d, a New Mouse and Human RAD51/RecA-Related Gene. *Genomics*. 1998. 49: 103-11.
88. Plon SE, Eccles DM, Easton D, Foulkes WD, Genuardi M, Greenblatt MS, et al. Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. *Human Mutation*. 2008. 29: 1282-91.
89. Rahman N, Seal S, Thompson D, Kelly P, Renwick A, Elliott A, et al. PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene. *Nature Genetics*. 2007. 39: 165-7.
90. Rebbeck TR, Mitra N, Wan F, Sinilnikova OM, Healey S, McGuffog L, et al. Association of Type and Location of *BRCA1* and *BRCA2* Mutations With Risk of Breast and Ovarian Cancer. *JAMA*. 2015. 313: 1347.
91. Reid S, Schindler D, Hanenberg H, Barker K, Hanks S, Kalb R, et al. Biallelic mutations in PALB2 cause Fanconi anemia subtype FA-N and predispose to childhood cancer. 2007. 39: 162-4.
92. Roy R, Chun J, Powell SN. BRCA1 and BRCA2: different roles in a common pathway of genome protection. *Nature Reviews Cancer*. 2011. 12: 68-78.
93. Ruffner H, Joazeiro CAP, Hemmati D, Hunter T, Verma IM. Cancer-predisposing mutations within the RING domain of BRCA1: Loss of ubiquitin protein ligase activity and protection from radiation hypersensitivity. *Proceedings of the National Academy of Sciences*. 2001. 98: 5134-9.
94. Saal, L. H. et al. The Sweden Cancerome Analysis Network - Breast (SCAN-B) Initiative: a large-scale multicenter infrastructure towards implementation of breast cancer genomic analyses in the clinical routine. *Genome medicine*. 2015. 7, 20.
95. Santarosa M, Viel A, Boiocchi M. Splice variant lacking the transactivation domain of the BRCA2 gene and mutations in the splice acceptor site of intron 2. *Genes Chromosomes Cancer*. 1999. 26: 381-2.
96. Santos C, Peixoto A, Rocha P, Pinto P, Bizarro S, Pinheiro M, et al. Pathogenicity Evaluation of BRCA1 and BRCA2 Unclassified Variants Identified in Portuguese Breast/Ovarian Cancer Families. *The Journal of Molecular Diagnostics*. 2014. 16: 324-34.
97. Sanz DJ, Acedo A, Infante M, Durán M, Pérez-Cabornero L, Esteban-Cardenosa E, Lastra E, Pagani F, Miner C, Velasco EA. A high proportion of DNA variants of BRCA1 and BRCA2 is associated with aberrant splicing in breast/ovarian cancer patients. *Cancer Res.*, 2010. 16: 1957-67.

98. Sibley CR, Blazquez L, Ule J. Lessons from non-canonical splicing. *Nat Rev Genet.* 2016. 17: 407-21.
99. Slater E, Langer P, Niemczyk E, Strauch K, Butler J, Habbe N, et al. PALB2 mutations in European familial pancreatic cancer families. *Clinical Genetics.* 2010. 78: 490-4.
100. Somyajit K, Saxena S, Babu S, Mishra A, Nagaraju G. Mammalian RAD51 paralogs protect nascent DNA at stalled forks and mediate replication restart. *Nucleic Acids Research.* 2015.
101. Spurdle AB, Couch FJ, Hogervorst FBL, Radice P, Sinilnikova OM, IARC Unclassified Genetic Variants Working Group. Prediction and assessment of splicing alterations: implications for clinical testing. *Hum Mutat.* 2008. 29: 1304-13.
102. Spurdle AB, Healey S, Devereau A, Hogervorst FBL, Monteiro ANA, Nathanson KL, et al. ENIGMA-Evidence-based network for the interpretation of germline mutant alleles: An international initiative to evaluate risk and clinical significance associated with sequence variation in BRCA1 and BRCA2 genes. *Human Mutation.* 2012. 33: 2-7.
103. Su H, Cheng Z, Huang J, Lin J, Copenhaver GP, Ma H, et al. Arabidopsis RAD51, RAD51C and XRCC3 proteins form a complex and facilitate RAD51 localization on chromosomes for meiotic recombination. Pawlowski WP, editor. *PLOS Genetics.* 2017. 13.
104. Suwaki N, Klare K and Tarsounas M. RAD51 paralogs: Roles in DNA damage signalling, recombinational repair and tumorigenesis. *Semin Cell Dev Biol.* 2011. 22:898-905.
105. Sy SMH, Huen MSY, Chen J. PALB2 is an integral component of the BRCA complex required for homologous recombination repair. *Proceedings of the National Academy of Sciences.* 2009. 106: 7155-60.
106. Taylor MRG, Špírek M, Chaurasiya KR, Ward JD, Carzaniga R, Yu X, et al. Rad51 Paralogs Remodel Pre-synaptic Rad51 Filaments to Stimulate Homologous Recombination. *Cell.* 2015.162: 271-86.
107. Tesoriero AA, Wong EM, Jenkins MA, Hopper JL, Brown MA, Chenevix-Trench G, et al. Molecular characterization and cancer risk associated with BRCA1 and BRCA2 splice site variants identified in multiple-case breast cancer families. *Hum Mutat.* 2005. 26: 495.
108. Thomassen M, Kruse TA, Jensen PK, Gerdes AM. A missense mutation in exon 13 in BRCA2, c.7235GA, results in skipping exon 13. *Genet Test.* 2006. 10: 116-20.
109. Thompson BA, Spurdle AB, Plazzer J-P, Greenblatt MS, Akagi K, Al-Mulla F, et al. Application of a 5-tiered scheme for standardized classification of 2,360 unique



- mismatch repair gene variants in the InSiGHT locus-specific database. *Nature Genetics*. 2013. 46: 107-15.
110. Tischkowitz M, Xia B. PALB2/FANCN: Recombining Cancer and Fanconi Anemia. *Cancer Researc*. 2010. 70: 7353-9.
111. Trapnell C, Pachter L and Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. 2009. *Oxford Academic*. 25: 1105–1111.
112. Traven A, Heierhorst J. SQ/TQ cluster domains: concentrated ATM/ATR kinase phosphorylation site regions in DNA-damage-response proteins. *BioEssays*. 2005. 27: 397-407.
113. Tung N, Lin NU, Kidd J, Allen BA, Singh N, Wenstrup RJ, et al. Frequency of Germline Mutations in 25 Cancer Susceptibility Genes in a Sequential Series of Patients With Breast Cancer. *Journal of Clinical Oncology*. 2016. 34: 1460-8.
114. Turunen JJ, Niemelä EH, Verma B, Frilander MJ. The significant other: splicing by the minor spliceosome. *Wiley Interdisciplinary Reviews: RNA*. 2013. 4: 61-76.
115. Vallée MP, Di Sera TL, Nix DA, Paquette AM, Parsons MT, Bell R, et al. Adding In Silico Assessment of Potential Splice Aberration to the Integrated Evaluation of BRCA Gene Unclassified Variants. *Human Mutation*. 2016. 37: 627-39.
116. Vega A, Campos B, Bressac-De-Paillerets B, Bond PM, Janin N, Douglas FS, et al. The R71G BRCA1 is a founder Spanish mutation and leads to aberrant splicing of the transcript. *Hum Mutat*. 2001. 17: 520-1.
117. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Kinzler KW. Cancer Genome Landscapes. *Science*. 2013. 339: 1546-58.
118. Vreeswijk MPG, Kraan JN, van der Klift HM, Vink GR, Cornelisse CJ, Wijnen JT, et al. Intronic variants in BRCA1 and BRCA2 that affect RNA splicing can be reliably selected by splice-site prediction programs. *Hum Mutat*. 2009. 30: 107-14.
119. Walker LC, Whiley PJ, Houdayer C, Hansen TVO, Vega A, Santamarina M, et al. Evaluation of a 5-tier scheme proposed for classification of sequence variants using bioinformatic and splicing assay data: inter-reviewer variability and promotion of minimum reporting guidelines. *Hum Mutat*. 2013. 34: 1424-31.
120. Walsh T, Casadei S, Lee MK, Pennil CC, Nord AS, Thornton AM, et al. Mutations in 12 genes for inherited ovarian, fallopian tube, and peritoneal carcinoma identified by massively parallel sequencing. *Proceedings of the National Academy of Sciences*. 2011. 108: 18032-7.
121. Wang Z, Burge CB. Splicing regulation: From a parts list of regulatory elements to an integrated splicing code. *RNA*. 2008. 14: 802-13.

122. Wang ET, Cody NAL, Jog S, Biancolella M, Wang TT, Treacy DJ, et al. Transcriptome-wide Regulation of Pre-mRNA Splicing and mRNA Localization by Muscleblind Proteins. *Cell*. 2012. 150: 710-24.
123. Whiley PJ, de la Hoya M, Thomassen M, Becker A, Brandao R, Pedersen IS, et al. Comparison of mRNA Splicing Assay Protocols across Multiple Laboratories: Recommendations for Best Practice in Standardized Clinical Testing. *Clinical Chemistry*. 2014. 60: 341-52.
124. Williams RS, Lee MS, Hau DD, Glover JNM. Structural basis of phosphopeptide recognition by the BRCT domain of BRCA1. *Nature Structural & Molecular Biology*. 2004. 11: 519-25.
125. Xia B, Dorsman JC, Ameziane N, de Vries Y, Rooimans MA, Sheng Q, et al. Fanconi anemia is associated with a defect in the BRCA2 partner PALB2. *Nature Genetics*. 2007. 39: 159-61.
126. Xia B, Sheng Q, Nakanishi K, Ohashi A, Wu J, Christ N, et al. Control of BRCA2 Cellular and Clinical Functions by a Nuclear Partner, PALB2. *Molecular Cell*. 2006. 22: 719-29.
127. Yang H, Jeffrey PD, Miller J, Kinnucan E, Sun Y, Thoma NH, et al. BRCA2 function in DNA binding and recombination from a BRCA2-DSS1-ssDNA structure. *Science*. 2002. 297: 1837-48.
128. Zhang F, Ma J, Wu J, Ye L, Cai H, Xia B, et al. PALB2 Links BRCA1 and BRCA2 in the DNA-Damage Response. *Current Biology*. 2009. 19: 524-9.



### **ARTÍCULOS PUBLICADOS O EN PREPARACIÓN**

- De la Hoya M, Soukarieh O, **López Perolio I**, et al. Combined genetic and splicing analysis of BRCA1 c.[594-2A>C; 641A>G] highlights the relevance of naturally occurring in-frame transcripts for developing disease gene variant classification algorithms. Hum Mol Genet. 2016; 25(11):2256-2268.
- Mara Colombo\*, **Irene López-Perolio\***, Huong D. Meeks, et al. When spliceogenic is not pathogenic: the case of the BRCA2 c.68-7T>A nucleotide change. 2016. Human Mutation (en proceso de revisión). \*These authors contributed equally to the work
- Rita D. Brandão, Klaas Mensaert, **Irene López-Perolio**, et al. Naturally occurring splicing events in tumour suppressor genes involved in breast/ovarian cancer (en preparación).
- Mads Thomassen, Thomas van Overeem Hansen, Mireia Menendez Vila, **Irene López-Perolio**, et al. Partial BRCA2 exon 3 skipping (en preparación).
- Gemma Montalbal, Eugenia Fraile-Bethencourt, **Irene López-Perolio et al.** Multi-centric analysis of BRCA2 splicing variant c.7976+5G>T identified in three unrelated Spanish families (en preparación).

### **PARTICIPACIONES EN CONGRESOS**

- **Irene López Perolio**, Ana Vega, Marta Santamariña et al. Splicing Alternativo y Mutaciones de Splicing en los genes de susceptibilidad al cáncer de mama y ovario BRCA1 y BRCA2: Aplicación de la PCR Digital. Premio al mejor poster presentado en el XXVIII congreso nacional de Genética Humana-AEGH, Palma de Mallorca, 2015.



ORIGINAL ARTICLE

# Combined genetic and splicing analysis of BRCA1 c.[594-2A>C; 641A>G] highlights the relevance of naturally occurring in-frame transcripts for developing disease gene variant classification algorithms

Miguel de la Hoya<sup>1,\*</sup>, Omar Soukarieh<sup>2</sup>, Irene López-Perolio<sup>1</sup>, Ana Vega<sup>3</sup>, Logan C. Walker<sup>4</sup>, Yvette van Ierland<sup>5</sup>, Diana Baralle<sup>6</sup>, Marta Santamariña<sup>7</sup>, Vanessa Lattimore<sup>4</sup>, Juul Wijnen<sup>5,8</sup>, Philip Whaley<sup>9</sup>, Ana Blanco<sup>3</sup>, Michela Raponi<sup>6</sup>, Jan Hauke<sup>10</sup>, Barbara Wappenschmidt<sup>10</sup>, Alexandra Becker<sup>10</sup>, Thomas V. O. Hansen<sup>11</sup>, Raquel Behar<sup>1</sup>, KConFaB Investigators<sup>12</sup>, Diether Niederacher<sup>13</sup>, Norbert Arnold<sup>14</sup>, Bernd Dworniczak<sup>15</sup>, Doris Steinemann<sup>16</sup>, Ulrike Faust<sup>17</sup>, Wendy Rubinstein<sup>18</sup>, Peter J. Hulick<sup>19</sup>, Claude Houdayer<sup>20</sup>, Sandrine M. Caputo<sup>21</sup>, Laurent Castera<sup>22</sup>, Tina Pesaran<sup>23</sup>, Elizabeth Chao<sup>23</sup>, Carole Brewer<sup>24</sup>, Melissa C. Southey<sup>25</sup>, Christi J. van Asperen<sup>5</sup>, Christian F. Singer<sup>26</sup>, Jan Sullivan<sup>27</sup>, Nicola Poplawski<sup>28,29</sup>, Phuong Mai<sup>30</sup>, Julian Peto<sup>31</sup>, Nichola Johnson<sup>32,33</sup>, Barbara Burwinkel<sup>34,35</sup>, Harald Surowy<sup>34,35</sup>, Stig E. Bojesen<sup>36,37,38</sup>, Henrik Flyger<sup>39</sup>, Annika Lindblom<sup>40</sup>, Sara Margolin<sup>41</sup>, Jenny Chang-Claude<sup>42,43</sup>, Anja Rudolph<sup>42</sup>, Paolo Radice<sup>44</sup>, Laura Galastri<sup>45</sup>, Janet E. Olson<sup>46</sup>, Emily Hallberg<sup>46</sup>, Graham G. Giles<sup>47,48</sup>, Roger L. Milne<sup>47,48</sup>, Irene L. Andrulis<sup>49,50</sup>, Gord Glendon<sup>49</sup>, Per Hall<sup>51</sup>, Kamila Czene<sup>51</sup>, Fiona Blows<sup>52</sup>, Mitul Shah<sup>52</sup>, Qin Wang<sup>53</sup>, Joe Dennis<sup>53</sup>, Kyriaki Michailidou<sup>53,54</sup>, Lesley McGuffog<sup>53</sup>, Manjeet K. Bolla<sup>53</sup>, Antonis C. Antoniou<sup>53</sup>, Douglas F. Easton<sup>53</sup>, Fergus J. Couch<sup>46,55</sup>, Sean Tavitgian<sup>56</sup>, Maaïke P. Vreeswijk<sup>5</sup>, Michael Parsons<sup>9</sup>, Huong D. Meeks<sup>56</sup>, Alexandra Martins<sup>2</sup>, David E. Goldgar<sup>57,†</sup> and Amanda B. Spurdle<sup>9,†,\*</sup>

<sup>†</sup>The authors wish it to be known that, in their opinion, the last two authors should be regarded as joint last Authors.

Received: November 23, 2015. Revised: February 10, 2016. Accepted: March 17, 2016

© The Author 2016. Published by Oxford University Press.

All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

<sup>1</sup>Molecular Oncology Laboratory, Instituto de Investigacion Sanitaria San Carlos (IdISSC), Hospital Clinico San Carlos, Madrid, Spain, <sup>2</sup>Inserm U1079-IRIB, University of Rouen, Normandy Centre for Genomic and Personalized Medicine, Rouen, France, <sup>3</sup>Fundacion Publica Galega de Medicina Xenómica-SERGAS Grupo de Medicina Xenómica-USC, IDIS, CIBERER, Santiago de Compostela 15706, Spain <sup>4</sup>Department of Pathology, University of Otago, Christchurch 8140, New Zealand, <sup>5</sup>Department of Clinical Genetics, Leiden University Medical Centre, Leiden 2300, The Netherlands, <sup>6</sup>Human Development and Health, Faculty of Medicine, University of Southampton, Southampton SO16 5YA, UK, <sup>7</sup>CIBERER, Grupo de Medicina Xenómica-USC, Universidade de Santiago de Compostela, Fundacion Galega de Medicina Xenómica (SERGAS), Santiago de Compostela 15706, Spain, <sup>8</sup>Department of Clinical Genetics, Leiden University Medical Centre, Leiden 2300, The Netherlands, <sup>9</sup>Department of Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, QLD 4006, Australia, <sup>10</sup>Center for Hereditary Breast and Ovarian Cancer, Center for Integrated Oncology (CIO), Medical Faculty, University Hospital Cologne, Cologne 50931, Germany, <sup>11</sup>Center for Genomic Medicine, Rigshospitalet, Copenhagen University Hospital, Copenhagen DK-2100, Denmark, <sup>12</sup>Peter MacCallum Cancer Center, University of Melbourne, Melbourne, VIC 3002, Australia, <sup>13</sup>Department of Gynaecology and Obstetrics, University Hospital Düsseldorf, Heinrich-Heine University Düsseldorf, Düsseldorf 40225, Germany, <sup>14</sup>Department of Gynaecology and Obstetrics, University Hospital of Schleswig-Holstein, Campus Kiel, Christian-Albrechts University Kiel, Kiel 24105, Germany, <sup>15</sup>Institute of Human Genetics, University of Münster, Münster 48149, Germany, <sup>16</sup>Institute of Human Genetics, Hannover Medical School, Hannover 30625, Germany, <sup>17</sup>Institute of Medical Genetics and Applied Genomics, University Hospital Tuebingen, Tuebingen 72076, Germany, <sup>18</sup>National Institutes of Health, Bethesda, MD 20892-2152, USA, <sup>19</sup>Center for Medical Genetics, NorthShore University Health System, University of Chicago Pritzker School of Medicine, Evanston, IL 60201, USA, <sup>20</sup>Service de Génétique, Department de Biologie des Tumeurs, Institut Curie and INSERM U830, Centre de Recherche de l'Institut Curie, Paris, and Université Paris Descartes, Sorbonne Paris Cité, Paris 75248, France, <sup>21</sup>Service de Génétique, Department de Biologie des Tumeurs, Institut Curie, Paris 75248, France, <sup>22</sup>Centre Francois Baclesse, Laboratoire de Biologie et de Genetique du Cancer, 14076 Caen, Paris 75248, France, <sup>23</sup>Ambry Genetics, Aliso Viejo, CA 92656, USA, <sup>24</sup>Department of Clinical Genetics, Royal Devon and Exeter Hospital, Exeter, UK, <sup>25</sup>Genetic Epidemiology Laboratory, Department of Pathology, The University of Melbourne, Melbourne, VIC 3010, Australia, <sup>26</sup>Department of Obstetrics and Gynaecology, Medical University of Vienna, Vienna, Austria, Waehringer Guertel 18-20, A 1090 Vienna, Austria, <sup>27</sup>Genetic Health Service NZ, South Island Hub, Christchurch Hospital, Christchurch 8140, New Zealand, <sup>28</sup>Adult Genetics Unit, South Australian Clinical Genetics Service, SA Pathology at the Women's and Children's Hospital, North Adelaide, SA 5067, Australia, <sup>29</sup>University Department of Paediatrics, University of Adelaide, North Terrace, Adelaide, SA 5000, Australia, <sup>30</sup>Clinical Genetics Branch, DCEG, NCI, NIH, Bethesda, MD, USA, <sup>31</sup>London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK, <sup>32</sup>Breakthrough Breast Cancer Research Centre, Institute of Cancer Research, London SW3 6JB, UK, <sup>33</sup>Division of Breast Cancer Research, Institute of Cancer Research, London SW3 6JB, UK, <sup>34</sup>Molecular Biology of Breast Cancer, Department of Gynecology and Obstetrics, University of Heidelberg, Heidelberg 69120, Germany, <sup>35</sup>Molecular Epidemiology Group, German Cancer Research Center, DKFZ, Heidelberg 69120, Germany, <sup>36</sup>Copenhagen General Population Study, Herlev and Gentofte Hospital, Copenhagen 2730, Denmark, <sup>37</sup>Department of Clinical Biochemistry, Herlev Hospital, Copenhagen University Hospital, Herlev 2730, Denmark, <sup>38</sup>Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen 1165, Denmark, <sup>39</sup>Department of Breast Surgery, Herlev and Gentofte Hospital, Copenhagen University Hospital, 2730 Denmark, <sup>40</sup>Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm SE-171 77, Sweden, <sup>41</sup>Department of Oncology Pathology, Karolinska Institutet, Stockholm SE-171 77, Sweden, <sup>42</sup>Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg 69120, Germany, <sup>43</sup>University Cancer Center Hamburg (UCCH), University Medical Center Hamburg-Eppendorf, Hamburg 20246, Germany, <sup>44</sup>Unit of "Molecular bases of genetic risk and genetic testing", Department of Preventive and Predictive Medicine, Fondazione IRCCS Istituto Nazionale dei Tumori, Milano 20139, Italy, <sup>45</sup>Associazione Volontari Italiani Sangue (AVIS) comunale di Milano, Milano 20139, Italy, <sup>46</sup>Department of Health Sciences Research, Mayo Clinic, Rochester, MN 55905, USA, <sup>47</sup>Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, University of Melbourne, VIC 3010, Australia, <sup>48</sup>Cancer Epidemiology Centre, Cancer Council Victoria, Melbourne, VIC 3004, Australia,

<sup>49</sup>Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, ON M5G 1X5, Canada,

<sup>50</sup>Department of Molecular Genetics, University of Toronto, M5B 1W8, Canada, <sup>51</sup>Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm SE-171 77, Sweden, <sup>52</sup>Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge CB1 8RN, UK, <sup>53</sup>Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge CB1 8RN, UK, <sup>54</sup>Department of Electron Microscopy/Molecular Pathology, The Cyprus Institute of Neurology and Genetics, 1683, Nicosia, Cyprus, <sup>55</sup>Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN 55905, USA, <sup>56</sup>Department of Oncological Sciences and <sup>57</sup>Department of Dermatology, Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, UT 84132, USA

\*To whom correspondence should be addressed at: QIMR Berghofer Medical Research Institute, 300 Herston Rd, Brisbane, QLD 4006, Australia. Tel: +617 3362 0371; Email: amanda.spurdle@qimr.edu.au (A.B.S.); Laboratorio de Oncología Molecular, Instituto de Investigación Sanitaria San Carlos (IdiSSC), Hospital Clínico San Carlos, c/Martin Lagos s/n, Madrid 28040, Spain. Tel: +34 913303348; Fax +34 913303544; Email: mhoya@hotmail.com (M.d.l.H)

## Abstract

A recent analysis using family history weighting and co-observation classification modeling indicated that BRCA1 c.594-2A > C (IVS9-2A > C), previously described to cause exon 10 skipping (a truncating alteration), displays characteristics inconsistent with those of a high risk pathogenic BRCA1 variant. We used large-scale genetic and clinical resources from the ENIGMA, CIMBA and BCAC consortia to assess pathogenicity of c.594-2A > C. The combined odds for causality considering case-control, segregation and breast tumor pathology information was  $3.23 \times 10^{-8}$ . Our data indicate that c.594-2A > C is always in cis with c.641A > G. The spliceogenic effect of c.[594-2A > C; 641A > G] was characterized using RNA analysis of human samples and splicing minigenes. As expected, c.[594-2A > C; 641A > G] caused exon 10 skipping, albeit not due to c.594-2A > C impairing the acceptor site but rather by c.641A > G modifying exon 10 splicing regulatory element(s). Multiple blood-based RNA assays indicated that the variant allele did not produce detectable levels of full-length transcripts, with a *per allele* BRCA1 expression profile composed of  $\approx 70$ – $80\%$  truncating transcripts, and  $\approx 20$ – $30\%$  of in-frame  $\Delta 9,10$  transcripts predicted to encode a BRCA1 protein with tumor suppression function.

We confirm that BRCA1c.[594-2A > C; 641A > G] should not be considered a high-risk pathogenic variant. Importantly, results from our detailed mRNA analysis suggest that BRCA-associated cancer risk is likely not markedly increased for individuals who carry a truncating variant in BRCA1 exons 9 or 10, or any other BRCA1 allele that permits 20–30% of tumor suppressor function. More generally, our findings highlight the importance of assessing naturally occurring alternative splicing for clinical evaluation of variants in disease-causing genes.

## Introduction

Sequence variants that alter the highly conserved intronic dinucleotides at splice donor and acceptor sites of high-risk disease predisposition genes are often assumed to be pathogenic, due to their high likelihood to alter RNA splicing. Although such variants will almost certainly lead to disruption of normal splicing patterns, the exact nature of the resulting alternate splicing patterns cannot be reliably predicted. Indeed, a standardized classification scheme recently developed for mismatch repair gene variants through consensus across multiple international sites (1) proposes that mRNA assay and/or clinical data are necessary to upgrade dinucleotide donor and acceptor variant classification from 'likely pathogenic' to 'pathogenic'.

The dinucleotide acceptor site variant BRCA1 c.594-2A > C (also known as IVS9-2A > C) has recently been reported associated with clinical characteristics inconsistent with a high risk of cancer expected for a pathogenic BRCA1 variant (2). Previous RNA analyses of carriers of BRCA1 c.594-2A > C indicate that this variant is associated with an aberrant mRNA profile (3,4), including production of exon 10 deleted out-of-frame transcripts. These observations indicate that the relationship between splicing aberrations and increased risk is not straightforward, and pose the question of which measures of mRNA transcript

dysregulation best reflect variant pathogenicity, considering recommendations already published by the ENIGMA Splicing Working Group (5). We undertook a study to assess level of risk associated with BRCA1 c.594-2A > C using segregation and large-scale case-control analysis, and detailed mRNA analyses correlating genotype with aberrant mRNA profiles.

## Results

### Genetic studies

Characteristics of BRCA1 c.594-2A > C variant carriers identified in BCAC, CIMBA and ENIGMA are detailed in Supplementary Material, Tables S1 and S2.

BRCA1 c.594-2A > C (rs80358033) was identified in 7/24 605 invasive breast cancer cases and 9/25 836 controls, when including only the 11 studies with at least one observation (Supplementary Material, Table S1[TQ1]). Standard case-control analysis yielded an odds ratio (OR) of 0.82 (95% CI 0.26–2.47), which was little different after adjustment for principle components (OR 0.83, 95% CI 0.41–2.24). However, some studies indicated that they had performed BRCA1/2 mutation screening of cases and may have excluded cases with pathogenic variants. Since BRCA1 c.594-2A > C



has generally been assumed to be pathogenic on the basis of its location at a splice acceptor site, this could create a bias due to preferential exclusion of c.594-2A > C carriers cases but not controls. After exclusion of four studies that did such genetic testing, we were left with 5/20 992 cases and 6/22 332 controls that carried the c.594-2A > C variant (see Supplementary Material, Table S1), yielding a revised OR of 0.87 (95% CI 0.26–2.86) after adjustment for principle components. The odds for causality based on carrier frequency and ages at diagnosis/interview in these cases and controls was  $7.3 \times 10^{-5}$  (equating to an odds against pathogenicity of 13 770:1). The case-control findings demonstrate that the BRCA1 c.594-2A > C variant is clearly not associated with a high risk of breast cancer, and is unlikely to be associated with even a moderate (~3- to 5-fold) risk of breast cancer. There were 15 BRCA1 c.594-2A > C carrier individuals from 13 families identified in the CIMBA dataset through genotyping with the iCOGS array. It was confirmed with the submitting sites that none of these individuals carried another pathogenic variant in BRCA1, and that eight of these families overlapped with those identified via ENIGMA while the proband for another family was also recruited into BCAC. Overall, information for segregation analysis was available for 14 probands from ENIGMA/CIMBA (Supplementary Material, Table S1), and breast tumor pathology information for 32 cases from ENIGMA, CIMBA or BCAC (Supplementary Material, Tables S1 and S2). The combined odds for causality based on segregation analysis, assuming BRCA1 age-specific risks of breast and ovarian cancer as estimated in the large study of Antoniou et al (6), was 0.10 (ranging from 0.02 to 6.85 for individual families). The breast tumor pathology features of variant carriers were not consistent with those found commonly for high-risk BRCA1 pathogenic variant carriers. The majority of tumors were estrogen receptor (ER) positive (25/32), and the odds for causality based on pathology information was  $4.98 \times 10^{-6}$  (200994:1 against causality).

After contacting the submitting centres and through re-investigation of original genetic test results, the BRCA1 exonic variant rs55680408 (c.641A > G, p.Asp214Gly) was confirmed to be present in all ENIGMA/CIMBA c.594-2A > C families included in the final analysis, and another 13 c.594-2A > C carriers identified by Ambry Genetics that were excluded from analysis due to lack of relevant clinical information. Specifically, review of genetic testing data by Ambry Genetics identified a total of 20 carriers of BRCA1 c.[594-2A > C; 641A > G] from >213 000 tests, including two siblings and a mother-daughter pair; there was clear evidence that the alleles were in cis from next-generation sequencing reads, and neither allele was observed alone in 2636 unrelated parent exomes. Further, all carrier individuals from BCAC were shown to share the same BRCA1 haplotype (data not shown). Based on the haplotype and genotype information, it was assumed that all BRCA1 c.594-2A > C carriers in the BCAC dataset were also carriers of c.641A > G (p.Asp214Gly). Considering case-control, segregation and pathology information, the combined odds for causality was  $3.61 \times 10^{-11}$  (i.e.  $2.77 \times 10^{10}$ :1 against causality). These results indicate that individuals carrying BRCA1c.[594-2A > C; 641A > G] (Supplementary Material, Fig. S1A) should be counselled as not having a high risk of BRCA1-associated disease.

## Splicing studies

### *Comprehensive characterization of BRCA1 alternative splicing landscape in c.[594-2A > C; 641A > G] carriers by capillary electrophoresis and high throughput RNA sequencing*

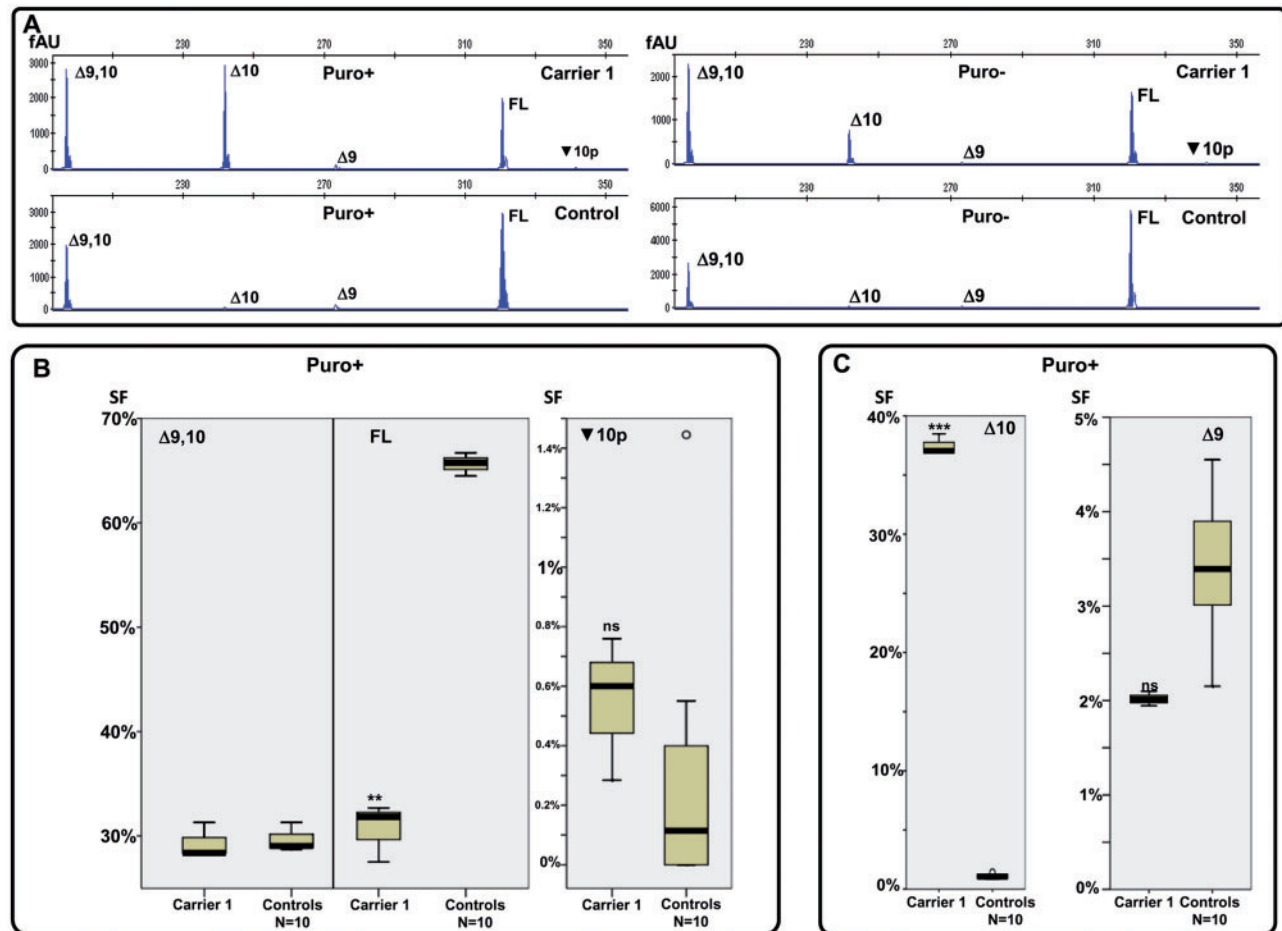
To search for a plausible biological mechanism explaining the lack of evidence for an increased cancer risk in BRCA1 c.[594-

2A > C; 641A > G] carriers, we first performed a comprehensive characterization of the BRCA1 alternative splicing landscape in the vicinity of exon 10. With this aim, we performed a series of capillary electrophoresis analyses on RNAs obtained from lymphoblastoid cell lines (LCLs) (see Materials and Methods). We have shown previously that this approach is highly sensitive, allowing comprehensive identification, characterization and semi-quantification of alternative splicing (4,7). Experiments performed with two combinations of forward and reverse primers located in exons 8 and 11 detected up to five different alternative splicing events both in LCLs from one c.[594-2A > C; 641A > G] carrier (Carrier 1) and healthy controls (Fig. 1), including three in-frame (full-length (FL),  $\Delta 9,10$ , and  $\nabla 10p$ ), and two out-of-frame ( $\Delta 9$  and  $\Delta 10$ ). All but  $\nabla 10p$  (r.594-21\_594-1ins) have been described previously as naturally occurring BRCA1 alternative splicing events in control samples (7). No c.[594-2A > C; 641A > G] specific events were identified. Overall, experiments conducted in the presence of puromycin (Puro+) experiments in Fig. 1 and Supplementary Material, Fig. S1) indicated that  $\Delta 9,10$  splicing fraction (SF) ( $\Delta 9,10^{SF}$ ) is similar in Carrier 1 and Controls ( $\approx 29\%$ ),  $\Delta 10^{SF}$  is considerably higher ( $\approx 38\%$  versus  $\approx 1\%$ ), and FL<sup>SF</sup> much lower ( $\approx 31\%$  versus  $\approx 66\%$ ).  $\Delta 9^{SF}$  ( $< 3\%$ ) and  $\nabla 10p^{SF}$  ( $< 1\%$ ) were rather minor alternative splicing events in all tested samples. As expected, Puro– experiments measured higher  $\Delta 9,10^{SF}$  in Carrier 1 than in Controls (Fig. 1A and Supplementary Material, Fig. S1) due to a drop in  $\Delta 10^{SF}$ , probably reflecting nonsense-mediated decay (NMD) degradation of out-of-frame  $\Delta 10$  transcripts.

Overall, findings were confirmed by comparable experiments performed by other contributing centers (Supplementary Material, Fig. S2), assaying up to eight individual variant allele carriers identified in four unrelated families and three different types of samples (LCLs, Leukocytes (LEUs), and fresh whole blood); there was similar  $\Delta 9,10^{SF}$  in Carriers and Controls (range 20–30% depending on specific protocols and/or sample used for experiments), and a significant increase of  $\Delta 10^{SF}$  (with corresponding decrease of FL<sup>SF</sup>) in Carriers. Complementary analyses performed in the subpopulation of BRCA1 $\Delta 11q$  transcripts were coincident, with similar ( $\Delta 9,10+\Delta 11q$ )<sup>SF</sup> in Carriers and Controls, and a significant increase of ( $\Delta 10+\Delta 11q$ )<sup>SF</sup> (with corresponding decrease of  $\Delta 11q^{SF}$ ) in Carriers (Supplementary Material, Fig. S2C). Incidentally, our data supports  $\nabla 10p$  as a naturally occurring BRCA1 alternative splicing event not previously reported, probably due to its very low SF. Capillary electrophoresis findings (in particular the lack of variant allele-specific transcripts, and the detection of  $\nabla 10p$  in Controls) were confirmed by RNA-seq experiments (Supplementary Material, Fig. S3).

### *Quantitative analyses combined with alternative splicing event specific biallelic expression analyses confirms that c.[594-2A > C; 641A > G] modifies the BRCA1 alternative splicing landscape, but not the overall BRCA1 expression level*

The comprehensive analysis of the BRCA1 alternative splicing landscape described above did not provide an obvious explanation for why c.[594-2A > C; 641A > G] carriers do not display features of a standard pathogenic BRCA1 variant. Yet, the absence of carrier-specific transcripts prompted us to speculate that it is perhaps the actual level of naturally occurring in-frame transcripts in variant allele carriers that may explain the genetic findings, in particular levels of FL and  $\Delta 9,10$  transcripts given that  $\nabla 10p$  transcript levels were very low. Since capillary electrophoresis is a semi-quantitative approach, we decided to perform further analyses with quantitative PCR (qPCR) and digital PCR (dPCR) that, overall, confirmed capillary electrophoresis findings (Fig. 2). qPCR



**Figure 1.** Capillary Electrophoresis analyses of BRCA1 alternative splicing landscape in LCLs from one BRCA1c.[594-2A>C; 641A>G] carrier (Carrier 1) and 10 Controls. **Panel A** shows representative examples of capillary electrophoresis analysis of RT-PCR products generated with the E8.1-E11p assay in LCLs treated (Puro+) or untreated (Puro-) with the NMD inhibitor puromycin. The fluorescence intensity of each peak (Y-axis) is expressed in arbitrary units (AU). The analyses detected the FL transcript, and up to four alternative splicing events, two in-frame ( $\Delta 9,10$  and  $\nabla 10p$ ) and two out-of-frame ( $\Delta 9$  and  $\Delta 10$ ). In these particular examples,  $\nabla 10p$  transcripts are detected only in Carrier 1, but we have detected  $\nabla 10p$  transcripts in Controls, as summarized in panel B. The presence of  $\nabla 10p$  in Controls has been further confirmed by RNAseq (see Supplementary Material, Fig. S3). The boxplots in **Panel B** (displaying low, Q1, median, Q3 and high values) show the SF of in-frame transcripts ( $\Delta 9,10$ , FL and  $\nabla 10p$ ) observed in Carrier 1 (three technical replicates) and 10 Controls. SF expressed as the % of the corresponding peak area to the  $\Sigma$  of all five peak areas detected by capillary electrophoresis. This particular experiment was performed with the E8.2-E11q.2 assay. Note that the  $\nabla 10p^{SF}$  is rather minor (<1%) regardless of the LCL tested. The FL<sup>SF</sup> was much lower in Carrier 1 than in Control samples. The boxplots in **Panel C** (displaying low, Q1, median, Q3 and high values) show the SF of out-of-frame transcripts ( $\Delta 9$  and  $\Delta 10$ ) observed in Carrier 1 (three technical replicates) and 10 Controls. The relative contribution of  $\Delta 10$  to the overall signal was much higher in Carrier 1 than in Control samples. Normal outliers (>1.5 interquartile range, IQR) display small circles. (– represents  $P \leq 0.01$ ) (\*\*\* represents  $P \leq 0.001$ ) (ns = non-significant).

absolute quantification of individual alternative splicing events in Carrier 1 estimated for  $\Delta 9,10^{SF}$  a value of  $19\% \pm 0.9$ , in the upper-limit of Controls (ranging from 4% to 17%; Fig. 2A, left), together with an obvious reduction of FL transcripts (Fig. 2A, right). In addition to  $\Delta 9,10$ , three other naturally occurring in-frame alternative splicing events involving exon 10 and/or nearby exon 11 have been described, namely  $\Delta 9_{11}$ ,  $\Delta 11$  and  $\Delta 11q$  (7). We used qPCR absolute quantification to estimate the SF of these alternative splicing events, detecting an increase of  $(\Delta 9,10 + \Delta 11q)^{SF}$  in Carrier 1 ( $9\% \pm 0.8$ ) if compared with Controls (average of 7%). No differences were observed with regard to  $\Delta 11^{SF}$  and  $\Delta 9_{11}^{SF}$  (Supplementary Material, Fig. S4). Similarly, dPCR analyses (Fig. 2B) revealed a modest increase of  $\Delta 9,10^{SF}$  in Carrier 1 ( $24\% \pm 0.9$ ) if compared with Controls (average of 17%), together with a 50% reduction of FL<sup>SF</sup> that is fully compatible with lack of FL transcripts arising from the variant allele.

Alternative splicing event specific reverse transcription and PCR amplification (RT-PCR) sequencing experiments

(Supplementary Material, Fig. S5) performed in carriers 3–5 (from one Dutch family) confirmed that  $\Delta 9,10$  expression is biallelic, whereas  $\Delta 10$  expression is essentially monoallelic (arising from the variant allele). Neither qPCR absolute quantification (Supplementary Material, Fig. S4A), nor biallelic expression analysis (Supplementary Material, Fig. S5B) suggested higher overall BRCA1 expression level in c.[594-2A>C; 641A>G] carriers. Yet, to further exclude this possibility we performed dPCR analyses of BRCA1 $\Delta 9,10$  and FL expression relative to BRCA2 (Supplementary Material, Fig. S6). The data indicated that  $\Delta 9,10$  relative expression level is similar in LCLs from Carrier 1 and Controls, while FL expression level shows a 50% reduction, again supporting that the variant allele is not producing FL transcripts.

Taken together, capillary electrophoresis analyses of RT-PCR products, RNA sequencing (RNAseq), qPCR, dPCR and alternative splicing event specific sequencing experiments supported a model in which the variant allele does not produce novel BRCA1

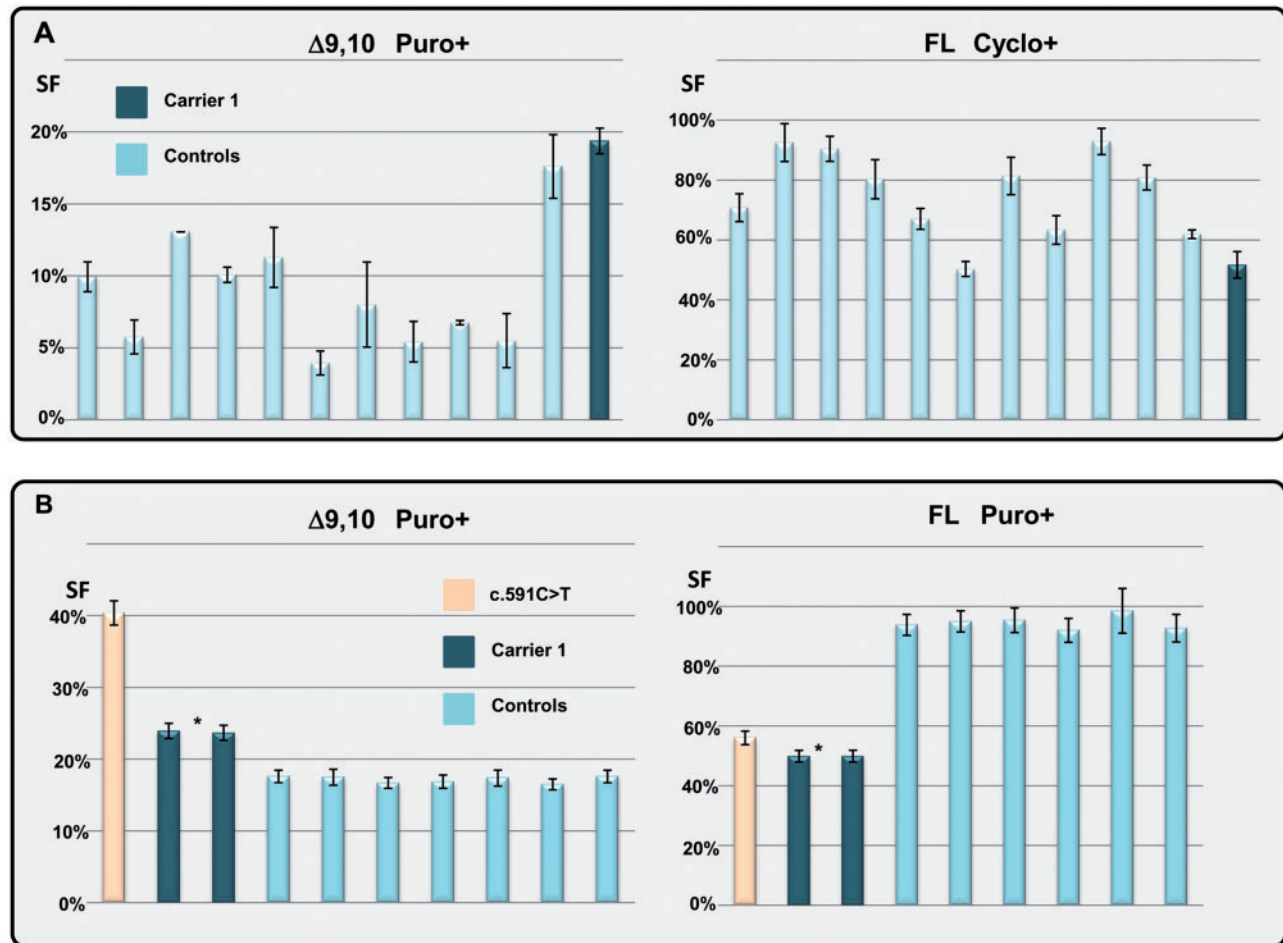


Figure 2. Quantification of major in-frame transcripts  $\Delta 9,10$  and FL in LCLs from one BRCA1c.[594-2A>C; 641A>G] carrier (Carrier 1) and Controls. Experiments were performed in LCLs treated with Puromycin (Puro+). Panel A displays  $\Delta 9,10^{\text{SF}}$  and FL<sup>SF</sup>, estimated as the ratio between the GAPDH normalized absolute numbers of  $\Delta 9,10$  (or FL) molecules and absolute number of all BRCA1 transcripts, as determined by qPCR analysis performed with standard curves (see Supplementary Material, Methods and Fig. S4). Standard deviation of 3 independent measures is shown. Panel B displays dPCR data measuring  $\Delta 9,10^{\text{SF}}$  and FL (inclusion of exons 9 and 10)<sup>SF</sup>, using exon23-24 junction as a proxy for overall BRCA1 expression level. The precision of each measure (as determined by the QuantStudio 3D Analysis Cloud Software) is indicated. Two technical replicates of Carrier 1 are shown. We included as positive control a LCL carrying the BRCA1 c.591C>T variant, known to increase  $\Delta 9,10^{\text{SF}}$ . The  $\Delta 9,10^{\text{SF}}$  in Carrier 1 was higher than in Controls (24% in two technical replicates of Carrier 1 versus an average of 17% in 7 control samples, Mann-Whitney U test;  $P = 0.028$  for difference between groups), but a 50% reduction of FL<sup>SF</sup> (50% in two technical replicates of Carrier 1 versus an average of 94% in six control samples, Mann-Whitney U test;  $P = 0.036$  for difference between groups).

transcripts, nor increases overall BRCA1 expression level, but rather substitutes FL transcripts (containing exons 9 and 10) with out-of-frame  $\Delta 10$  transcripts, such that the contribution of in-frame  $\Delta 9,10$  transcripts to the overall expression level is similar or slightly higher (see Fig. 2, Supplementary Material, Fig. S2) to that observed in wild-type (WT) alleles. Of note, according to our data the overall model is also probably true in the subset of BRCA1A11q transcripts (see Supplementary Material, Figs S2C and S4B). According to this model, BRCA1 $\Delta 9$  (out-of-frame) and BRCA1  $\nabla 10$ p (in-frame) contribution to the overall expression level are very low both in variant and WT alleles (see Fig. 2B and C), and hence irrelevant to explain the lack of risk observed in variant allele carriers.

#### Splicing reporter minigene analyses reveal that c.641A>G is causing exon 10 skipping in c.[594-2A>C; 641A>G] carriers

We also performed minigene assay experiments to dissect the contribution of the individual variants c.594-2A>C and c.641A>G to the splicing pattern observed in variant allele

carriers. Experiments were performed with two minigene assays (pCAS2-BRCA1-Exon10 and pB1). A schematic representation of these reporter minigenes is shown in Figure 3 (8). pCAS2-BRCA1-Exon10 and pB1 experiments performed in HeLa cells, as well as pB1 experiments performed in breast (MCF7 and HBL100) and ovarian (IGROV-1) cell lines, revealed that both c.594-2A>C and c.641A>G impair normal exon 10 splicing, albeit with different outcomes (Fig. 3). pCAS2-BRCA1-Exon10 c.594-2A>C and pB1 c.594-2A>C predominantly produced  $\nabla 10$ p transcripts, but also a minor amount of  $\Delta 10$  transcripts (Fig. 3A), a finding confirming previous pSPL3-BRCA1-Exon10 experiments performed in COS-7 cells (9). In contrast, pCAS2-BRCA1-Exon10 c.641A>G and pB1 c.641A>G mostly produced  $\Delta 10$  but no detectable  $\nabla 10$ p. The finding that c.641A>G causes exon 10 skipping albeit being located outside the splice site, suggests that this variant disturbs the regulation of exon 10 splicing, probably by destroying splicing enhancer elements and/or by creating splicing silencer elements, a hypothesis supported by an *in silico* analysis based on ESRseq scores (Supplementary Material, Fig. S7A). The presence of regulatory mechanisms

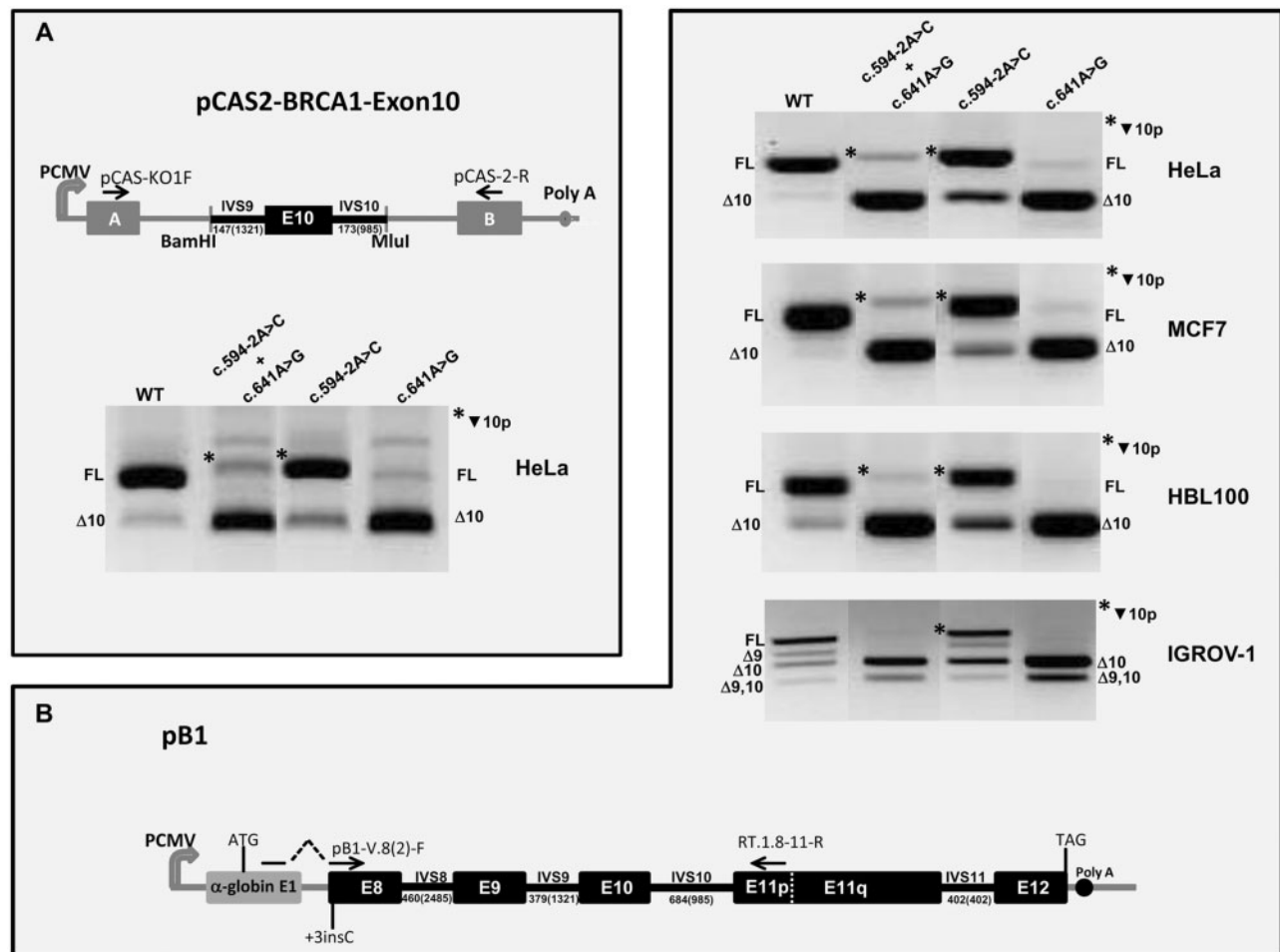
underlying BRCA1 exon 10 splicing was further supported by small interfering RNA experiments performed in MDA-MB231 cells showing that endogenous BRCA1 depends on Tra2- $\beta$  for exon 10 inclusion (Supplementary Material, Fig. S7B). Double mutant pCAS2-BRCA1-Exon10 c.[594-2A>C; 641 A>G] and pB1 c.[594-2A>C; 641 A>G] experiments mimicking the variant allele observed *in vivo* produced detectable levels of both  $\Delta 10$  and  $\nabla 10p$ , with  $\Delta 10$  being the predominant outcome in all cell lines tested (Fig. 3).

## Discussion

In this study, we have demonstrated that c.[594-2A>C; 641A>G] carriers (but not necessarily carriers of a potential BRCA1 allele in which c.594-2A>C is not linked to c.641A>G) should not be considered at high-risk of developing BRCA1-associated cancers. The finding is remarkable, since the variant

allele causes exon 10 skipping, a frame-shift alteration. In addition, we propose a plausible biological mechanism underlying the finding, the so-called BRCA1  $\Delta 9,10$  rescue model, and we show the relevance of the findings for developing disease gene variant classification algorithms.

The first study addressing the spliceogenic impact of BRCA1c.594-2A>C demonstrated an association with exon 10 skipping (3), supporting the initial pathogenic classification by Myriad Genetics (2). Here we confirm exon 10 skipping in c.594-2A>C carriers, and we show that contrary to expectations this splicing alteration is not driven by c.594-2A>C, but rather by the linked variant c.641A>G. Further, we show that the variant allele does not produce FL transcripts, nor other in-frame transcripts apart from normal levels of  $\Delta 9,10$  and residual levels of  $\nabla 10p$  transcripts. These findings lead us to conclude that  $\Delta 9,10$  transcripts arising from the variant allele confer sufficient tumor suppressor activity *in vivo* to compensate for the lack of FL



**Figure 3.** Analysis of BRCA1 c.594-2A>C and c.641A>G variants with splicing reporter minigene assays. The figure shows schematic non-scale representations of the splicing reporter minigenes pCAS2-BRCA1-exon10 (panel A) and pB1 (panel B) used for splicing assays. Minigenes were constructed as described under Supplementary Material, Methods. PCMV indicates the cytomegalovirus promoter, boxes represent exons and lines in between indicate introns. BRCA1 sequences are highlighted in black. Arrows represent primers used in RT-PCR reactions. With the exception of pB1 BRCA1 intron 11 (402 nt-long FL IVS11), minigenes harbor partial segments of BRCA1 introns. For comparative purposes, the size in nucleotides of each segment is shown together with the size corresponding to the endogenous FL BRCA1 introns shown in brackets. As indicated, pB1 carries an additional cytosine (+3insC) in exon 8 to keep the ORF with  $\alpha$ -globin exon 1 (8). Splicing assays were performed by analyzing the splicing pattern of WT and mutant minigenes (c.594-2A>C, c.641A>G and c.[594-2A>C; 641A>G]) transiently expressed in human cells (HeLa, COS-7, MCF7, HBL100 or IGROV-1) as described under Supplementary Methods. The images show RT-PCR products separated in ethidium bromide-stained agarose gels. FL, full-length;  $\Delta 9$ , exon 9 skipping;  $\Delta 10$ , exon 10 skipping;  $\Delta 9,10$ , skipping of both exons 9 and 10; asterisk, retention of 21 intronic nucleotides immediately upstream exon 10 ( $\nabla 10p$ ). One can note that: (i) the relative level of alternatively spliced pB1(WT) transcripts is higher in IGROV-1 than in HeLa, MCF-7 or HBL100 cells, and (ii) the predominant alternative splicing event of pB1(WT) in these cell lines is  $\Delta 10$ , whereas that of endogenous wild-type BRCA1 in blood related samples is  $\Delta 9,10$  (Fig. 4 and Supplementary Material, Figs 1 and 2).



transcripts. To be more precise, the combined genetic and splicing data lead us to formulate a  $\Delta 9,10$  rescue model in which BRCA1 alleles with an associated  $\Delta 9,10^{SF}$  of  $\approx 20\text{--}30\%$  (as measured in blood related samples) confer tumor suppressor haplo-sufficiency (Fig. 4). The actual value is probably closer to 20% than to 30% (according both to qPCR and dPCR estimations in Carrier 1, and to capillary electrophoresis estimations in Carriers 2–8), but at any rate is very similar to that observed in control samples. The finding that  $\Delta 9,10$  is a predominant alternative splicing event not only in blood derived samples but also in clinically relevant tissues such as breast and ovary (Supplementary Material, Fig. S8) is critical to support our rescue model for both breast and ovarian cancer. Indeed, family history of breast and/or ovarian cancer is a key criterion for genetic testing for most participating ENIGMA and CIMBA sites, and segregation analysis modelled both breast and ovarian cancer risk, providing no indication that BRCA1 c.594-2A > C (IVS9-2A > C) could be associated with increased ovarian cancer risk only. Further, similar to our findings reported for breast cancer, case-control data from a parallel study by the Ovarian Cancer Association Consortium does not support an association with ovarian cancer risk, with BRCA1 c.594-2A > C identified in 2/16 121 cases and 4/26 167 controls (OCAC, unpublished data). Note that the BRCA1  $\Delta 9,10$  rescue model predicts lack of breast and ovarian cancer risk not only for BRCA1 variants causing exon 10 skipping (or exon 9 skipping), but to any loss-of-function mutation in exons 9 or 10 (nonsense or frame shift mutations), provided that the mutant allele produces normal levels of  $\Delta 9,10$  transcripts (Fig. 4).

Evidently, the BRCA1  $\Delta 9,10$  rescue model presumes that  $\Delta 9,10$  transcripts encode a protein isoform (BRCA1<sup>p.Gly183\_Lys223del</sup>) that has tumor suppressor activity. To our knowledge, this BRCA1 isoform (lacking only 41 out of 1863 amino acid residues) has not been detected *in vivo*, nor functionally characterized *in vitro*, but tumor suppressor activity is fully compatible with structural considerations: 1) the 41 missing residues are unlikely to affect protein folding, since they are embedded in an intrinsically disordered protein region spanning amino acids 170–1649 (10); 2) BRCA1<sup>p.Gly183\_Lys223del</sup> includes all known functional domains/residues critical for tumor suppression, including the RING domain (spanning amino acids 2–103) that mediates binding to BARD1, an obligated heterodimer partner *in vivo* (11). Interestingly, BRCA1<sup>p.Gly183\_Lys223del</sup> lacks some residues critical for E3 ligase activity (12), a BRCA1 function that appears to be dispensable for tumor suppression (13,14). Yet, the most compelling argument supporting BRCA1<sup>p.Gly183\_Lys223del</sup> tumor suppressor activity stems from combined genetic and splicing analyses of BRCA1 c.591C > T (rs1799965). This variant, also not associated with the high risk of cancer expected for a pathogenic BRCA1 variant (current odds for causality of  $8.50 \times 10^{-16}$  based on segregation and pathology information, ENIGMA unpublished data), expresses mostly  $\Delta 9,10$  transcripts, a significant proportion of out-of-frame  $\Delta 9$  transcripts, and very few FL transcripts (15), strongly pointing to BRCA1<sup>p.Gly183\_Lys223del</sup> as a protein with tumor suppressor function. As far as we know, the only cancer predisposition gene for which a similar alternative splicing rescue model has been proposed is the tumor suppressor *adenomatous polyposis coli* (APC) gene, albeit in this case loss of function variants in the alternatively spliced region of APC exon 9 are not associated with lack of risk, but with a milder phenotype, termed attenuated familial adenomatous polyposis (16).

The BRCA1 $\Delta 9,10$  rescue model highlights the often neglected relevance of naturally occurring alternative splicing in the clinical arena, and has obvious implications for variant classification

algorithms. The ENIGMA consortium has developed and documented criteria for the five-tier classification of BRCA1/2 genetic variants based on qualitative and quantitative information (<http://www.enigmaconsortium.org/>, last accessed March 30, 2016). According to these rules, and consistent with those proposed by InSiGHT for Mismatch Repair gene variants (1), BRCA1/2 variants considered extremely likely to alter splicing based on position (typically  $IVS \pm 1$  or  $IVS \pm 2$ ) were initially all considered Class-4 (likely pathogenic) if untested for splicing alterations. However, the findings presented in this study have been pivotal to support amendment to these classification criteria, specifying need for particular caution in interpreting variants in instances where  $\Delta 9,10$  (or other known naturally occurring in-frame alternative splicing events) might rescue gene functionality (see Supplementary Material, Table S3). Hence, we also recommend caution in interpreting coding sequence variants that lead to premature termination codons in BRCA1 exons 9 and 10. This conservative stance is consistent with recent American College of Medical Genetics (ACMG) guidelines (17), which recommend considering the presence of alternative gene transcripts, understanding which are biologically relevant, and in which tissues the products are expressed. Thus, caution should be exercised when interpreting the impact of truncating variants confined to only a subset of transcripts, given the presence of other protein isoforms.

Of note, our results have additional implications unrelated to alternative splicing. More precisely, our study suggests that BRCA1 tumor suppressor activity tolerates a substantial reduction in expression level *in vivo*. Indeed, results shown in Figure 4 indicate that a BRCA1 allele producing as much as  $\approx 70\text{--}80\%$  of transcript encoding tumor suppressor deficient protein (as measured in blood-related samples) may not necessarily confer high-risk of developing cancer. This observation supports the conservative viewpoint of the ENIGMA consortium that, in the absence of other information, a variant can be considered pathogenic due to an effect on mRNA integrity if it only produces transcripts carrying a premature stop codon or an in-frame deletion disrupting known functional domain(s), as determined by semi-quantitative or quantitative methods.

In brief, there are several broad messages arising from the present study. Our results confirm that mRNA and genetic studies are warranted to inform the clinical significance of sequence alterations at the highly conserved intronic dinucleotides of splice donor and acceptor sites, and highlight the need to consider both variant haplotype and alternative splicing events in the design and interpretation of assays assessing the functional consequences of variants of uncertain clinical significance. We have also shown that comprehensive understanding of alternative splicing, paired with clinical genetic studies, is critical to understand the clinical consequences of complex splicing profiles observed for certain spliceogenic variants. Lastly, we provide a baseline hypothesis for future investigation and interpretation of other likely spliceogenic BRCA1/2 variants, a hypothesis that has implications for informing standards for generic variant classification guidelines.

## Materials and methods

### Genotyping and Sample Sets

We undertook screening of BRCA1 c.594-2A > C by direct genotyping, as part of the iCOGS experiment detailed elsewhere (18,19). This study included genotype and pathology results from breast cancer cases and controls participating in the

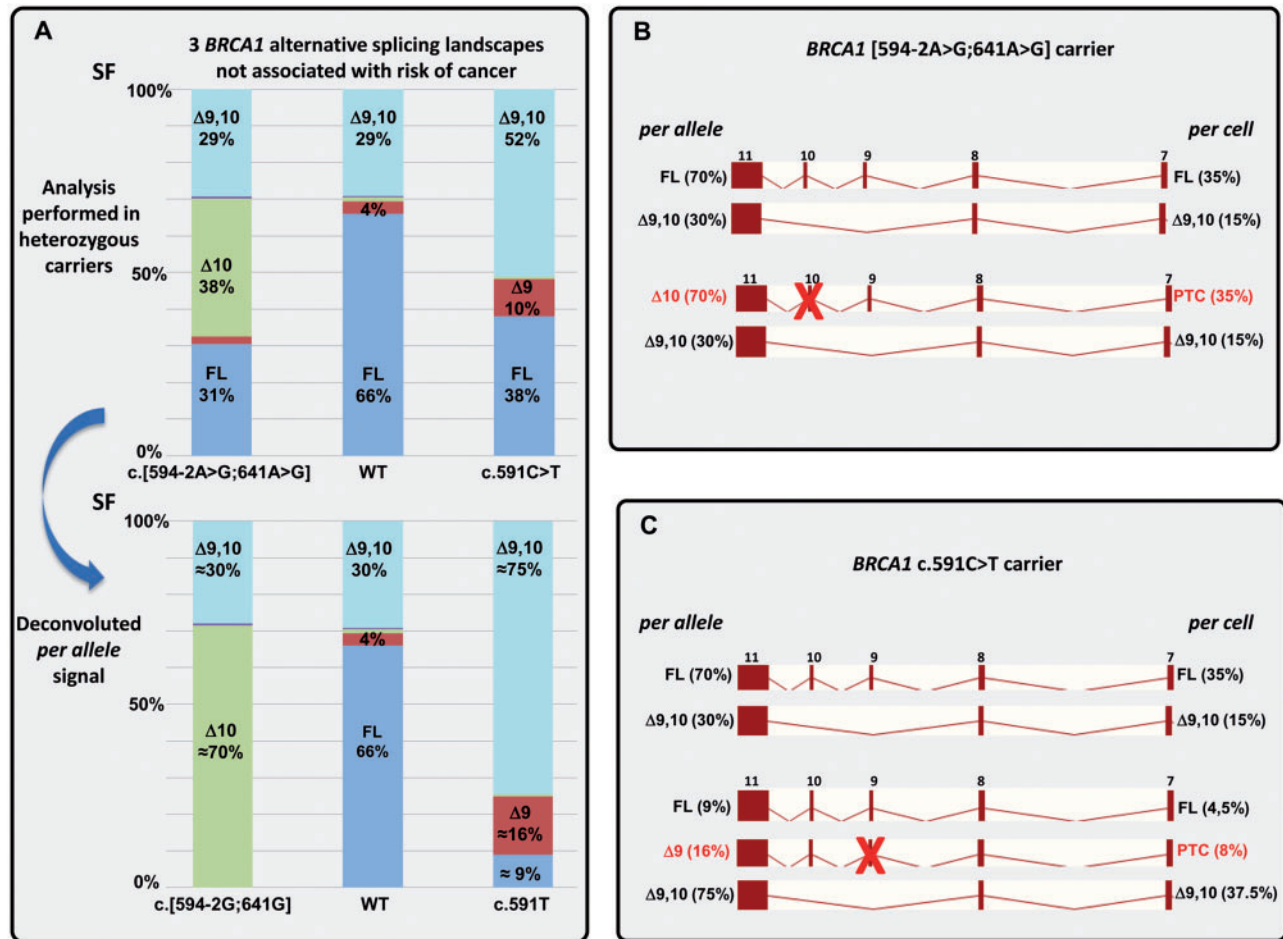


Figure 4. Combined genetic and splicing analyses of *BRCA1* c.[594-2A>G; 641A>G] and *BRCA1* c.591C>T supports a *BRCA1*Δ9,10 rescue model with far-reaching clinical implications. Panel A (top) shows the SF of five alternative splicing events detected by capillary electrophoresis analysis of RT-PCR products generated with the E8.2-E11q.2 assay (Puro+ experiments, 36 cycle PCRs, see Fig.1 and Supplementary Material, Fig. 1 for further details). As shown, this description of the *BRCA1* alternative splicing landscape in the vicinity of exon 10 is different in healthy control samples, c.[594-2A>G; 641A>G] carriers, and c.591C>T carriers. Yet, we show in the present study that none of these three *BRCA1* splicing landscapes is associated with high risk of developing *BRCA1*-related cancers. The chart displays SFs that, in carriers, represent a combined signal from the variant allele and the accompanying WT allele. Panel A (bottom). Deduced per allele SFs are shown. Assuming that SFs arising from the accompanying WT allele equal to the average SFs observed in 10 Control samples (as shown in the central chart bar), we deconvoluted the SFs corresponding to c.[594-2G; c.641G] (left chart bar) and c.591T (right chart bar) alleles. Panel B. The cartoon represents the relative per allele (100% equals to the overall expression level arising from one individual allele) and per cell (100% equals to the overall expression arising from a diploid genome) expression (*BRCA1* exons 7–11) in a c.[594-2G; c.641G] carrier, inferred from capillary EP analyses shown in Panel A. For simplicity, only FL and Δ9,10 transcripts are shown, albeit Δ9 and Δ10p transcripts account for ≈5% of the per cell expression. Truncating (out-of-frame) events are highlighted with a red cross. The analysis suggests that expressing up to ≈35% of *BRCA1* PTC-NMD transcripts (per diploid genome) is not associated with high-risk of developing cancer. The analysis suggests as well that a *BRCA1* allele expressing up to ≈70% (per allele) *BRCA1* PTC-NMD transcripts is not associated with high-risk of developing cancer (a relevant finding in the context of the two-hit model). Panel C. The cartoon represents the relative per allele (100% equals to the overall expression level arising from one individual allele) and per cell (100% equals to the overall expression arising from a diploid genome) expression (*BRCA1* exons 7–11) in a c.591C>T carrier, inferred from capillary EP analyses shown in Panel A. For simplicity, only FL, Δ9,10 and Δ9 (variant allele) are shown, albeit Δ9 (wt allele), Δ10 (wt and variant allele), and Δ10p (wt and variant allele) transcripts account for ≈5% of the per cell expression. The data strongly suggests that *BRCA1*Δ9,10 transcripts, representing up to 51% (per diploid genome) and up to 71% (per allele) of the overall *BRCA1* expression code for a *BRCA1* protein with tumor suppressor activity. The model displayed in this figure is intended to illustrate the most relevant findings of our study. Yet, some limitations should be highlighted. First, the model assumes (based on 36-cycle PCR capillary EP data) that Δ9,10SF in Controls and c.[594-2A>G; 641A>G] carriers is ≈29%, while other experiments suggests that the actual value is probably lower in both instances (Fig. 2, Supplementary Material, Fig. 2), albeit slightly increased in Carriers versus Controls. The model has been elaborated with data obtained in LCLs, not in clinically relevant tissues such as breast or ovarian. PTC, premature termination codon.

Breast Cancer Association Consortium (BCAC; <http://apps.ccge.medschl.cam.ac.uk/consortia/bcac/>, last accessed March 30, 2016), and from carriers of *BRCA1* assumed pathogenic variants participating in the Consortium of Investigators of Modifiers of *BRCA1/2* (CIMBA; <http://apps.ccge.medschl.cam.ac.uk/consortia/cimba/>, last accessed March 30, 2016). In addition, via the Evidence-based Network for Investigating Germline Mutant Alleles (ENIGMA, <http://enigmaconsortium.org/>, last accessed

March 30, 2016 (20)), we identified probands recruited through familial cancer clinics who were found to be positive for *BRCA1* c.594-2A>G via clinical genetic testing. All study participants were enrolled into national or regional studies under ethically approved protocols.

Information was recorded for all variant carriers regarding cancer status, age at diagnosis/interview, breast tumor pathology (grade and ER, Progesterone Receptor (PR), and Herceptin-2 (HER2)

status), and also pedigree and segregation information where available. For carriers identified through CIMBA and ENIGMA, the genotype for exonic variant c.641A > G (p.Asp214Gly) was sought from the original clinical testing report.

The BCAC dataset included 53 354 breast cancer cases and 49 720 controls and documented age at diagnosis/interview from 45 studies, detailed in (18). The denominator reduced to 24 605 cases and 25 836 controls when including only invasive breast cancer cases and controls from the 11 studies with at least one observation (Supplementary Material, Table S1). These 11 studies included only individuals of European ancestry, and four (MCBS, MBCCSG, KARBAC, OFBCR) had undergone testing for germline *BRCA1/2* pathogenic variants (4–100% of samples, depending on the BCAC study), including two of four studies which sampled cases on the basis of reported family history or presence of bilateral disease.

The CIMBA dataset included 11 105 female *BRCA1* pathogenic variant carriers aged  $\geq 18$  years from 46 studies in CIMBA recruited through cancer genetics clinics. There were 4845 females without report of cancer, 4713 breast cancer cases, 933 ovarian cancer cases, and 614 individuals reporting both breast and ovarian cancer.

By contact with submitters and examination of clinical information, it was established that 11 of the 15 CIMBA probands overlapped with individuals included in the ENIGMA dataset, and one of proband was also a participant in BCAC site (Supplementary Material, Table S2). Only non-overlapping data was included in multifactorial likelihood analysis.

## Statistical methods

We evaluated the effect of the *BRCA1* c.594-2A > C variant on breast cancer risk in BCAC, using logistic regression models with adjustment for censoring age and population structure, based on six principal components which defined any residual population sub-structure. Censoring age was defined as age at breast cancer diagnosis, or age at last interview/follow-up. Only case-control studies in which the variant was observed at least once were included in the analysis.

In order to place case-control data into the same likelihood ratio (LR) framework as the other lines of evidence used for multifactorial likelihood analysis (21,22), we compared the likelihood of the distribution of *BRCA1* c.594-2A > C variant carriers among cases and controls under the hypothesis that the variant has the same age-specific relative risks as the 'average' pathogenic *BRCA1* mutation compared to that under the hypothesis that it is not associated with any increased breast cancer risk. Specifically, we used the age at diagnosis of cases and age at interview for controls together with the relative risks of breast cancer estimated from case series unselected for family history (6) to calculate the probability that each individual carrying *BRCA1* c.594-2A > C in the sample is a *BRCA1* pathogenic variant carrier given their affected status and age. Under the hypothesis that *BRCA1* c.594-2A > C is a benign variant and does not confer increased breast cancer risk, we calculated the probability of the distribution of cases and controls among *BRCA1* c.594-2A > C carriers as a simple binomial probability with  $p$  = proportion of cases in the sample. These two likelihoods were then compared to derive the appropriate LR.

Bayes scores for segregation were derived as described previously (23), and pathology LRs were applied as indicated in Spurdle et al. (24). The segregation scores, pathology LRs and case-control LRs are mutually independent and were combined to derive a combined odds for causality as described previously (21).

We used the program Phase 2.0 (25) to estimate the most likely haplotypes of the BCAC cases and controls based on 29 variants in the region within and surrounding the *BRCA1* locus, in order to examine if all c.594-2A > C variant carriers were observed on the same haplotypic background. Variants used for phasing were those submitted by ENIGMA for inclusion on the iCOGS chip design, the most common of which were rs8176258, rs1799967, rs1799950, rs4986852 and rs1799966.

## mRNA analysis methods

### Nomenclature

We use as reference sequences to describe *BRCA1* genetic variants the GenBank reference sequences U14680.1 (cDNA) and NC\_000017.11 (genomic). When referring to *BRCA1* exons, we use exon numbering according to U14680.1. To characterize the *BRCA1* alternative splicing landscape in c.[594-2A > C; 641A > G] carriers (sometimes referred throughout the text as *variant allele* carriers), we performed different RNA splicing analyses at the immediate vicinity of *BRCA1* exon 10 (defined as the gene region spanning exons 8–11). Since our methodology do not allow analysis of complete transcripts (from 5'-end to poly(A) tail), we refer throughout the text to alternative splicing event containing transcripts, or alternative splicing events, rather than to alternative splicing transcripts or RNA isoforms (7). For the very same reason, FL refers throughout the text to *BRCA1* exons 9- and 10-containing transcripts (transcripts containing the exons9/10 junction defined in the GenBank reference sequence U14680.1), and not necessarily to the complete 5711nt mRNA described in U14680.1. We have designated alternative splicing events by combining U14680.1 exon numbering with the following symbols:  $\Delta$  (exon skipping),  $\blacktriangledown$  (intron retention),  $p$  (proximal, or 5') and  $q$  (distal, or 3').

### RNA analysis of human samples

Up to seven contributing laboratories (sites 1–7) performed RNA splicing analyses with various methodologies, including fluorescent RT-PCR followed by capillary electrophoresis, real-time qPCR, dPCR, Sanger sequencing and RNAseq (see Supplementary Material, Methods for further details). Experiments were performed in RNAs extracted from LCLs, short-term (3–6 days) cultured LEU or fresh peripheral blood. RNAs were derived from eight individual c.[594-2A > C; 641A > G] carriers (hereafter referred as Carriers 1–8) identified in four unrelated families from Australia (Carrier 1, LCL), Germany (Carrier 2, LCL), The Netherlands (Carriers 3–7, LEUs), and France (Carrier 8, peripheral blood) and healthy controls. We conducted several experiments designed to characterize the *BRCA1* alternative splicing landscape observed in *variant carriers*. We used as quantitative description the SF, defined here as the contribution of individual alternative splicing events to the overall *BRCA1* expression level (expressed as a percentage). As proxies for overall expression level, we used the  $\Sigma$  of all peak areas detected (capillary electrophoresis), or the signal obtained with a TaqMan assay recognizing the *BRCA1* exons 23 and 24 junction (dPCR). The latter was selected since both *BRCA1* exons 23 and 24 are likely constitutive exons (7). Note that SF is a relative measure between signals arising from the same locus (in this case *BRCA1*), so that it is neither directly related to the actual expression level on individual splicing events, nor with the overall expression level from that locus. It is formally possible that increments in the SF of one particular alternative splicing event correlate with actual reductions in the expression level of



that splicing event. For that reason, we determine the absolute expression level of individual alternative splicing events by qPCR with standard curves (see Supplementary Material, Methods for further details), and we performed relative expression analyses by dPCR, using as a reference a TaqMan assay recognizing the BRCA2 exons 26–27 junction. When indicated, we used as a positive control RNA extracted from LCLs carrying the BRCA1 variant c.591C > T [p.=(Cys197Cys)], known to increase  $\Delta 9^{SF}$  and  $\Delta 9,10^{SF}$  (15). Many experiments were performed in parallel with cultured cells treated/untreated with a nonsense mediated mRNA decay pathway (NMD) inhibitor, either Puromycin (Puro+/- experiments) or Cycloheximide (Cyclo+/- experiments). RNA from Carrier 8 was directly extracted from fresh peripheral blood. Biallelic expression was assessed by alternative splicing event specific RT-PCR followed by Sanger sequencing through rs1060915 (an informative exonic SNP located at BRCA1 exon 13), using primers and protocols previously described (15). In addition, we searched for BRCA1 tissue specific alternative splicing landscape in clinically relevant samples by comparing RNAs extracted from healthy control fresh peripheral blood, a pool of 10 healthy breast tissues (enriched normal epithelial areas selected by a pathologist) adjacent to breast tumor samples and commercial RNAs from healthy breast and ovarian human tissues. Experiments were performed by capillary electrophoresis of RT-PCR products, and by dPCR. Depending on the contributing laboratories, different RNA isolation and cDNA synthesis approaches were used (see Supplementary Material, Methods for further details).

#### Minigene splicing assays

To dissect the contribution of the individual BRCA1 variants c.594-2A > C and c.641A > G to the splicing alteration observed in c.[594-2A > C; 641A > G] carriers, we performed splicing assays with two different types of reported minigenes: pCAS2-BRCA1-Exon10 and pB1 (a minigene spanning BRCA1 exons 8–12). See Supplementary Material, Methods and Figure 3 for further details.

#### RNA interference experiments

To identify splicing regulatory proteins involved in BRCA1 exon 10 splicing, we performed a series of RNA interference experiments knocking down diverse splicing regulatory factors (hnRNPA1, Tra2 $\beta$ , SF2/ASF and SC35). Experiments were performed in the breast cancer cell line MDAMD231 (see Supplementary Material, Methods for further details).

### Supplementary material

Supplementary Material is available at HMG online.

### Acknowledgements

We thank all the families and individuals that participated in this research. We thank Paul Pharoah and the Ovarian Cancer Association Consortium for providing summary information on the frequency of the BRCA1 c.594-2A > C variant in ovarian cancer cases and controls.

We acknowledge the contributions of Georgia Chenevix-Trench to CIMBA and the kConFab resource, and additional study-specific acknowledgements as noted below.

NZBCS: Anne Smith, Bridget Robinson, Caroline Lintott, John Pearson, Yen Phung George Wiggins and the family members for their valuable contributions.

*The Netherlands Consortium:* FransHogervorst for assistance in data collation, Dr M. Olderode-Berends from the University Medical Centre Groningen and Dr E. van Riel from University Medical Centre Utrecht for providing blood samples for RNA studies, and Elsa Bik from the Leiden University Medical Centre, The Netherlands for excellent technical assistance.

*kConFab:* Heather Thorne, Eveline Niedermayr, all the kConFab research nurses and staff, the heads and staff of the Family Cancer Clinics, and the Clinical Follow Up Study for their contributions to this resource.

*GC-HBOC:* Dieter Schäfer, Center Frankfurt, for providing DNA samples and Juliane Köhler for excellent technical assistance.

*Northshore:* Tina Selkirk for assistance in data collation.

*French Consortium:* the French oncogeneticists, and the UNICANCER Genetic Group (UGG) lead by Dr Catherine Noguès.

*BCFR-AU:* Maggie Angelakos, Judi Maskiell, Gillian Dite, Helen Tsimiklis.

*MUV:* Daniela Muhr for assistance in data collation.

*BBCS:* Eileen Williams, Elaine Ryder-Mills, Kara Sargus.

*BSUCH:* Peter Bugert, Medical Faculty Mannheim.

*CGPS:* Staff and participants of the Copenhagen General Population Study. For the excellent technical assistance: Dorthe Uldall Andersen, Maria Birna Arnadottir, Anne Bank, Dorthe Kjeldgård Hansen. The Danish Breast Cancer Group (DBCG) is acknowledged for the tumor information. The Danish Cancer Biobank is acknowledged for providing infrastructure for the collection of blood samples for the cases.

*MARIE:* Alina Vrieling, Katharina Buck, MuhabbetCelik, Ursula Eilber and Sabine Behrens.

*MBCSG:* SiranoushManoukian, Bernard Peissel, Jacopo Azzollini and Fernando Ravagnani of the Fondazione IRCCS IstitutoNazionaleTumori (INT), Milan, Italy; BernandoBonanni, Monica Barile and Irene Feroce of the IstitutoEuropeo di Oncologia (IEO), Milan, Italy; and the personnel of the Cogentech Cancer Genetic Test Laboratory, Milan, Italy.

*OFBCR:* Teresa Selander, Nayana Weerasooriya.

*SEARCH:* Marie Mack.

*COGS:* Paul Pharoah, Andrew Berchuck (OCAC), Georgia Chenevix-Trench, Ken Offit (CIMBA), Alison M. Dunning, Andrew Lee, Ed Dicks, Craig Luccarini, the staff of the Centre for Genetic Epidemiology Laboratory, Javier Benitez, Anna Gonzalez-Neira, the staff of the CNIO genotyping unit, Jacques Simard, Daniel C. Tessier, Francois Bacot, Daniel Vincent, Sylvie La Boissière, Frederic Robidoux, the staff of the McGill University and Génome Québec Innovation Centre, Sune F. Nielsen, Borge G. Nordestgaard, the staff of the Copenhagen DNA laboratory, Julie M. Cunningham, Sharon A. Windebank, Christopher A. Hilker, Jeffrey Meyer, the staff of Mayo Clinic Genotyping Core Facility.

*AM and INSERM:* Dr Sophie Krieger for contributing with DNA samples, Professor Thierry Frébourg for providing patient's samples for RNA analysis, and AurélieDrouet and Gaia Castelain for technical assistance.

*Conflicts of interest:* Tina Pesaran and Elizabeth Chao are paid employees of Ambry Genetics. All other authors have declared no conflicts of interest.

### Funding

The research described was supported by Spanish Instituto de Salud Carlos III funding, an initiative of the Spanish Ministry of Economy and Innovation partially supported by European Regional Development FEDER Funds [PI12/00539 and PI15/00059].



to M.d.H., PI13/02030 to A.V.]; the French Ministry of Higher Education and Research [to O.S.]; the University of Otago, Mackenzie Charitable Foundation, Maria Lupton, and .Health Research Council of New Zealand [to L.W.]; UK Higher Education Funding Council Senior Fellowship Scheme, the University of Southampton [to D.B.]; Cancer research UK [to D.B., M.R.]; FamilienHede Nielsen Foundation fund [to T.V.O.H.]; Cancer Research-UK Senior Cancer Research Fellowship [to A.C.A.]; National Institute of Health [CA128978 and CA11616 to F.J.C.]; an NIH specialized program of research excellence in breast cancer to the Mayo Clinic [P50 CA116201 to F.J.C.]; and the US Breast Cancer Research Foundation [to F.J.C.]; translational grant from the French National Cancer Institute and Direction Générale de l'Offre des Soins (INCa-DGOS AAP/CFB/CI) and a grant from the French North-West Canceropole (CNO) [to A.M.]; The Cancer Council Queensland [APP1086286 to A.B.S.]; the NHMRC Senior Research Fellowship Scheme [ID 1061779 to A.B.S.]; NHMRC Project grant scheme [ID 1010719 to A.B.S.].

Additional infrastructure support to consortium members is as follows: kConFab infrastructure has been supported by funding from the National Breast Cancer Foundation, National Health and Medical Research Council, the Queensland Cancer Fund, the Cancer Councils of New South Wales, Victoria, Tasmania and South Australia, and the Cancer Foundation of Western Australia [to kConFab, and the kConFab Clinical Follow-up study].

The German Consortium of Hereditary Breast and Ovarian Cancer (GC-HBOC) is supported by the German Cancer Aid (grant no 109076, Rita K. Schmutzler) and by the Center for Molecular Medicine Cologne (CMMC)

The French consortium is supported by the French National Cancer Institute.

EMBRACE is supported by Cancer Research UK Grants C1287/A10118 and C1287/A11990.

BCFR was supported by grant UM1 CA164920 from the National Cancer Institute. The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centers in the Breast Cancer Family Registry (BCFR), nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government or the BCFR.

The BBCS is funded by Cancer Research UK and Breakthrough Breast Cancer (recently merged with Breast Cancer Campaign forming Breast Cancer Now) and acknowledges NHS funding to the NIHR Biomedical Research Centre, and the National Cancer Research Network (NCRN).

The BSUCH study was supported by the Dietmar-Hopp Foundation, the Helmholtz Society and the German Cancer Research Center (DKFZ).

The CGPS was supported by the Chief Physician Johan Boserup and LiseBoserup Fund, the Danish Medical Research Council and Herlev Hospital

KARBAC was supported financially through the regional agreement on medical training and clinical research (ALF) between Stockholm City Council and KarolinskaInstitutet, and from the Stockholm Cancer Foundation and the Swedish Cancer Society.

KARBAC was supported financially through the regional agreement on medical training and clinical research (ALF) between Stockholm City Council and KarolinskaInstitutet, and from the Stockholm Cancer Foundation and the Swedish Cancer Society.

The MARIE study was supported by the Deutsche Krebshilfe.V. [70-2892-BR I], the Hamburg Cancer Society, the

German Cancer Research Center and the Federal Ministry of Education and Research (BMBF) Germany [01KH0402].

The MARIE study was supported by the Deutsche Krebshilfe.V. [70-2892-BR I], the Hamburg Cancer Society, the German Cancer Research Center and the Federal Ministry of Education and Research (BMBF) Germany [01KH0402].

MBCSG is supported by grants from the Italian Association for Cancer Research (AIRC) and by funds from the Italian citizens who allocated the 5/1000 share of their tax payment according to Italian laws in support of the Fondazione IRCCS Istituto Nazionale Tumori.

The MCBCS was supported by the NIH grant CA128978 and a Specialized Program of Research Excellence (SPORE) in Breast Cancer [CA116201], the Breast Cancer Research Foundation and a generous gift from the David F. and Margaret T. Grohne Family Foundation and the Ting Tsung and Wei Fong Chao Foundation.

MCCS cohort recruitment was funded by VicHealth and Cancer Council Victoria. The MCCS was further supported by Australian NHMRC grants 209057, 251553 and 504711 and by infrastructure provided by Cancer Council Victoria.

OFBCR was supported by grant UM1 CA164920 from the National Cancer Institute. The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centers in the Breast Cancer Family Registry (BCFR), nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government or the BCFR.

The pKARMA study was supported by Märit and Hans Rausing's Initiative Against Breast Cancer, and the Swedish Medical Research Council.

SEARCH was supported by grants CRUK A490/A11021, C490/A16561.

Funding for the iCOGS infrastructure came from: the European Community's Seventh Framework Programme under grant agreement no. 223175 (HEALTH-F2-2009-223175) (COGS), Cancer Research UK (C1287/A10118, C1287/A 10710, C12292/A11174, C1281/A12014, C5047/A8384, C5047/A15007, C5047/A10692), the National Institutes of Health (CA128978) and Post-Cancer GWAS initiative (1U19 CA148537, 1U19 CA148065 and 1U19 CA148112 - the GAME-ON initiative), the Department of Defence (W81XWH-10-1-0341), the Canadian Institutes of Health Research (CIHR) for the CIHR Team in Familial Risks of Breast Cancer, Komen Foundation for the Cure, the Breast Cancer Research Foundation and the Ovarian Cancer Research Fund.

CIMBA data management was supported by Cancer Research-UK grant C12292/A11174 and C1287/A10118.

BCAC is funded by Cancer Research UK [C1287/A10118, C1287/A12014] and by the European Community's Seventh Framework Programme under grant agreement number 223175 (grant number HEALTH-F2-2009-223175) (COGS).

## References

1. Thompson, B.A., Spurdle, A.B., Plazzer, J.P., Greenblatt, M.S., Akagi, K., Al-Mulla, F., Bapat, B., Bernstein, I., Capella, G., den Dunnen, J.T. et al. (2014) Application of a 5-tiered scheme for standardized classification of 2,360 unique mismatch repair gene variants in the InSiGHT locus-specific database. *Nat. Genet.*, **46**, 107–115.
2. Rosenthal, E.T., Bowles, K.R., Pruss, D., van Kan, A., Vail, P.J., McElroy, H. and Wenstrup, R.J. (2015) Exceptions to the rule: case studies in the prediction of pathogenicity for genetic variants in hereditary cancer genes. *Clin. Genet.*, **88**, 533–541.

3. Tesoriero, A.A., Wong, E.M., Jenkins, M.A., Hopper, J.L., Brown, M.A., Chenevix-Trench, G., Spurdle, A.B., Southey, M.C. and kConFab. (2005) Molecular characterization and cancer risk associated with BRCA1 and BRCA2 splice site variants identified in multiple-case breast cancer families. *Hum. Mutat.*, **26**, 495.
4. Whiley, P.J., de la Hoya, M., Thomassen, M., Becker, A., Brandao, R., Pedersen, I.S., Montagna, M., Menendez, M., Quiles, F., Gutierrez-Enriquez, S. et al. (2014) Comparison of mRNA splicing assay protocols across multiple laboratories: recommendations for best practice in standardized clinical testing. *Clin. Chem.*, **60**, 341–352.
5. Walker, L.C., Whiley, P.J., Houdayer, C., Hansen, T.V., Vega, A., Santamarina, M., Blanco, A., Fachal, L., Southey, M.C., Lafferty, A. et al. (2013) Evaluation of a 5-tier scheme proposed for classification of sequence variants using bioinformatic and splicing assay data: inter-reviewer variability and promotion of minimum reporting guidelines. *Hum. Mutat.*, **34**, 1424–1431.
6. Antoniou, A., Pharoah, P.D., Narod, S., Risch, H.A., Eyfjord, J.E., Hopper, J.L., Loman, N., Olsson, H., Johannsson, O., Borg, A. et al. (2003) Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case Series unselected for family history: a combined analysis of 22 studies. *Am. J. Hum. Genet.*, **72**, 1117–1130.
7. Colombo, M., Blok, M.J., Whiley, P., Santamarina, M., Gutierrez-Enriquez, S., Romero, A., Garre, P., Becker, A., Smith, L.D., De Vecchi, G. et al. (2014) Comprehensive annotation of splice junctions supports pervasive alternative splicing at the BRCA1 locus: a report from the ENIGMA consortium. *Hum. Mol. Genet.*, **23**, 3666–3680.
8. Raponi, M., Douglas, A.G., Tammara, C., Wilson, D.I. and Baralle, D. (2012) Evolutionary constraint helps unmask a splicing regulatory region in BRCA1 exon 11. *PLoS One*, **7**, e37255.
9. Steffensen, A.Y., Dandanell, M., Jonson, L., Ejlersen, B., Gerdes, A.M., Nielsen, F.C. and Hansen, T. (2014) Functional characterization of BRCA1 gene variants by mini-gene splicing assay. *Eur. J. Hum. Genet.*, **22**, 1362–1368.
10. Mark, W.Y., Liao, J.C., Lu, Y., Ayed, A., Laister, R., Szymczyna, B., Chakrabarty, A. and Arrowsmith, C.H. (2005) Characterization of segments from the central region of BRCA1: an intrinsically disordered scaffold for multiple protein-protein and protein-DNA interactions? *J. Mol. Biol.*, **345**, 275–287.
11. Clark, S.L., Rodriguez, A.M., Snyder, R.R., Hankins, G.D. and Boehning, D. (2012) Structure-function of the tumor suppressor BRCA1. *Comput. Struct. Biotechnol. J.*, **1**, pii: e201204005.
12. Starita, L.M., Young, D.L., Islam, M., Kitzman, J.O., Gullingsrud, J., Hause, R.J., Fowler, D.M., Parvin, J.D., Shendure, J. and Fields, S. (2015) Massively parallel functional analysis of BRCA1 RING Domain Variants. *Genetics*, **200**, 413–422.
13. Shakya, R., Reid, L.J., Reczek, C.R., Cole, F., Egli, D., Lin, C.S., deRooij, D.G., Hirsch, S., Ravi, K., Hicks, J.B. et al. (2011) BRCA1 tumor suppression depends on BRCT phosphoprotein binding, but not its E3 ligase activity. *Science*, **334**, 525–528.
14. Savage, K.I. and Harkin, D.P. (2015) BRCA1, a 'complex' protein involved in the maintenance of genomic stability. *FEBS J.*, **282**, 630–646.
15. Dosil, V., Tosar, A., Canadas, C., Perez-Segura, P., Diaz-Rubio, E., Caldes, T. and de la Hoya, M. (2010) Alternative splicing and molecular characterization of splice site variants: BRCA1 c.591C>T as a case study. *Clin. Chem.*, **56**, 53–61.
16. Nieuwenhuis, M.H. and Vasen, H.F. (2007) Correlations between mutation site in APC and phenotype of familial adenomatous polyposis (FAP): a review of the literature. *Crit. Rev. Oncol. Hematol.*, **61**, 153–161.
17. Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E. et al. (2015) Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.*, **17**, 405–424.
18. Michailidou, K., Hall, P., Gonzalez-Neira, A., Ghoussaini, M., Dennis, J., Milne, R.L., Schmidt, M.K., Chang-Claude, J., Bojesen, S.E., Bolla, M.K. et al. (2013) Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat. Genet.*, **45**, 353–361.
19. Couch, F.J., Wang, X., McGuffog, L., Lee, A., Olswold, C., Kuchenbaecker, K.B., Soucy, P., Fredericksen, Z., Barrowdale, D., Dennis, J. et al. (2013) Genome-wide association study in BRCA1 mutation carriers identifies novel loci associated with breast and ovarian cancer risk. *PLoS Genet.*, **9**, e1003212.
20. Spurdle, A.B., Healey, S., Devereau, A., Hogervorst, F.B., Monteiro, A.N., Nathanson, K.L., Radice, P., Stoppa-Lyonnet, D., Tavtigian, S., Wappenschmidt, B. et al. (2012) ENIGMA—evidence-based network for the interpretation of germline mutant alleles: an international initiative to evaluate risk and clinical significance associated with sequence variation in BRCA1 and BRCA2 genes. *Hum. Mutat.*, **33**, 2–7.
21. Goldgar, D.E., Easton, D.F., Deffenbaugh, A.M., Monteiro, A.N., Tavtigian, S.V., Couch, F.J. and Breast Cancer Information Core (BIC) Steering Committee. (2004) Integrated evaluation of DNA sequence variants of unknown clinical significance: application to BRCA1 and BRCA2. *Am. J. Hum. Genet.*, **75**, 535–544.
22. Goldgar, D.E., Easton, D.F., Byrnes, G.B., Spurdle, A.B., Iversen, E.S., Greenblatt, M.S. and the IARC Unclassified Genetic Variants Working Group. (2008) Genetic evidence and integration of various data sources for classifying uncertain variants into a single model. *Hum. Mutat.*, **29**, 1265–1272.
23. Thompson, D., Easton, D.F. and Goldgar, D.E. (2003) A full-likelihood method for the evaluation of causality of sequence variants from family data. *Am. J. Hum. Genet.*, **73**, 652–655.
24. Spurdle, A.B., Couch, F.J., Parsons, M.T., McGuffog, L., Barrowdale, D., Bolla, M.K., Wang, Q., Healey, S., Schmutzler, R., Wappenschmidt, B. et al. (2014) Refined histopathological predictors of BRCA1 and BRCA2 mutation status: a large-scale analysis of breast cancer characteristics from the BCAC, CIMBA, and ENIGMA consortia. *Breast Cancer Res.*, **16**, 3419.
25. Delaneau, O., Marchini, J. and Zagury, J.F. (2012) A linear complexity phasing method for thousands of genomes. *Nat. Methods*, **9**, 179–181.